DOCUMENT RESUME

ED 266 166 TM 860 110

AUTHOR	Arter, Judith A.; Estes, Gary D.
TITLE	Item Banking for Local Test Development: Practitioner's Handbook.
INSTITUTION	Northwest Regional Educational Lab., Portland, OR. Assessment and Evaluation Program.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	Nov 85
CONTRACT	400-83-0005-P-15
NOTE	99p.; Appendices A and B contain small type. For a related document, see TM 860 111.
PUB TYPE	Guides - Non-Classroom Use (055)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Adaptive Testing; *Computer Assisted Testing; Computer Software; Curriculum; Elementary Secondary Education; Flow Charts; *Item Banks; Microcomputers; Resource Allocation; Scores; Surveys; *Teacher Made Tests; *Test Construction; Test Format; Testing Problems; *Testing Programs; Test Items; Test Use

ABSTRACT

This handbook is intended for persons who might develop or use an item bank to support their testing program. An item bank is defined as a "large collection of distinguishable test items," with "large" explained as meaning that the number of items available is greater than the number to be used in any one test. The first section of the handbook provides guidance as to the types of testing options which might be most appropriate for different testing purposes, resources, and local testing climate. The other two major sections deal with two item banking options: (1) accessing an existing item bank; and (2) developing one's own item bank. Because this handbook is intended as a practical guide, each section has three major parts: (1) a list of questions to guide users through decisions to be made on each topic; (2) assistance with answering questions; (3) examples to illustrate the various concepts presented. Extensive appendices include: (1) an item bank survey summary; (2) a summary of item banking software for microcomputers; (3) a summary of general purpose software for microcomputers which could be used for item banking; (4) item bank design questions; (5) test selection forms; and (6) sample classification scheme for reading items. (LMO)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

ED266166

Item Banking for Local Test Development

Practitioner's Handbook

By

Judith A. Arter
Gary D. Estes
Northwest Regional Educational Laboratory
Evaluation and Assessment
300 S.W. Sixth Avenue
Portland, Oregon 97204

TM 860 110

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

November 1985

This work was performed pursuant to Contract No. 400-83-0005, P-15 with the National Institute of Education, U.S. Department of Education. It does not, however, necessarily reflect the views of that agency and no official endorsement by the Institute should be inferred.

INTRODUCTION

This handbook is intended for persons who might develop or use an item bank to support their testing program. For purposes of this handbook, an item bank will be defined as a large collection of distinguishable test items (Estes and Arter, 1983). "Large" means that the number of items available is greater than the number to be used in any one test. "Collection" implies that the items, whether developed by the user or someone else, are kept together in some retrievable form. "Distinguishable" means that the items carry some information that permits the test constructor to select precisely those items he or she wants to use for each test.

This definition does not require that items be stored in a computer. For certain applications a computer would definitely be recommended; but manual systems are sometimes more appropriate. This definition also does not require that any particular type of information be stored along with items (for example Rasch calibrations) although, again for certain applications, calibrations would be useful.

We've purposefully made our definition nonrestrictive because our goal is to explore possible item banking uses, not to describe any particular system or application. We take the general philosophy that no single testing approach, including item banking, is appropriate for all test users. Therefore, the first section of the handbook provides some guidance as to the types of testing options which might be most appropriate for different testing purposes, resources, and local testing climate. Testing options could include commercially published tests, locally developed tests, an item bank with or without a computer or calibrations, a commercially available item bank, or informal tests. The other two major sections of the handbook deal with two item banking options: accessing an existing item bank, and developing one's own item bank.

This handbook has three major parts. Since item banking or customized testing may not be the best testing choice for some, the first section assists the user in thinking through what testing options might be most feasible in his or her testing program. The second section assists those who feel that customized testing (using another's item bank) is the right testing option. The third section presents considerations for those who wish to pursue their own item bank as a viable alternative.

Because this handbook is intended as a practical guide, each section has three major parts. White sheets provide a list of questions to guide users through decisions to be made on each topic. Blue sheets provide assistance with answering the questions. Yellow sheets provide examples to illustrate the various concepts presented. Each list of questions is presented as a flow chart, so that users can enter the handbook at any point and consider questions and concerns of relevance to them.

In this handbook we emphasize testing in cognitive domains in elementary, junior high, and high school. Although our examples and terminology reflect this major purpose, many of the concepts can be applied to other testing situations--testing for employment selection, for certifying professional competency, and for examination of other performance domains than the cognitive.

TABLE OF CONTENTS

	Page
WHICH TESTING OPTIONS SHOULD BE PURSUED?	1
Testing Options	4
Testing Purposes	8
Test—Curriculum Match	10
Practicalities	13
Sociopolitical Considerations	15
Need for Rigorous Item Development	16
What Testing Options Might Best Meet My Needs?	17
USING ITEM BANKS MAINTAINED BY OTHERS	19
Writing Test Specifications	21
Auxiliary Services	23
Resources	24
Selecting an Item Bank	25
Sample Test Specifications	27
Customized Testing Examples	32
DEVELOPING YOUR OWN ITEM BANK	34
What Resources are Available?	37
Item Banking Types	38
General Types of Item Banking Systems	40
Item Types, Numbers and Classification Scheme	45
How will Tests be Developed from the Bank?	47
Auxiliary Functions	49
Acquisition and Management of Items	50
Examples of Item Banks	52
References	57

TABLE OF CONTENTS --Cont'd

Page

APPENDICES

Appendix A -- Item Bank Survey Summary 58

Appendix B -- Summary of Item Banking Software for Microcomputers 63

Appendix C -- Summary of General Purpose Software for Microcomputers Which Could Be Used for Item Banking 71

Appendix D -- Item Bank Design Questions 73

Appendix E -- Test Selection Forms 81

Appendix F -- Sample Classification Scheme for Reading Items 86

LIST OF FIGURES

Page

Flowchart 1 -- Which Testing Options Should Be Pursued? 1

Table 1 -- Deciding on a Testing Option 18

Flowchart 2 -- Using Item Banks Developed By Others . . 20

Flowchart 3 -- Developing Your Own Item Bank 34

Table 2 -- Item Banking/Types 38

LIST OF EXAMPLES

Page

Test Specifications 27

Using Item Banks Developed By Others 32

Item Banks 52

LIST OF FORMS

Page

Short Form 1 — Purposes for Testing 9

Short Form 2 -- Need for Test-Curriculum Match 12

Short Form 3 — Testing Practicalities 14

Short Form 4 -- Sociopolitical Concerns 15

Short Form 5 -- Need for Rigorous Item Development . . . 16

Short Form 6 -- Resources Required for Using Someone Else's Item Bank 24

Short Form 7 -- Selecting an Item Bank 25

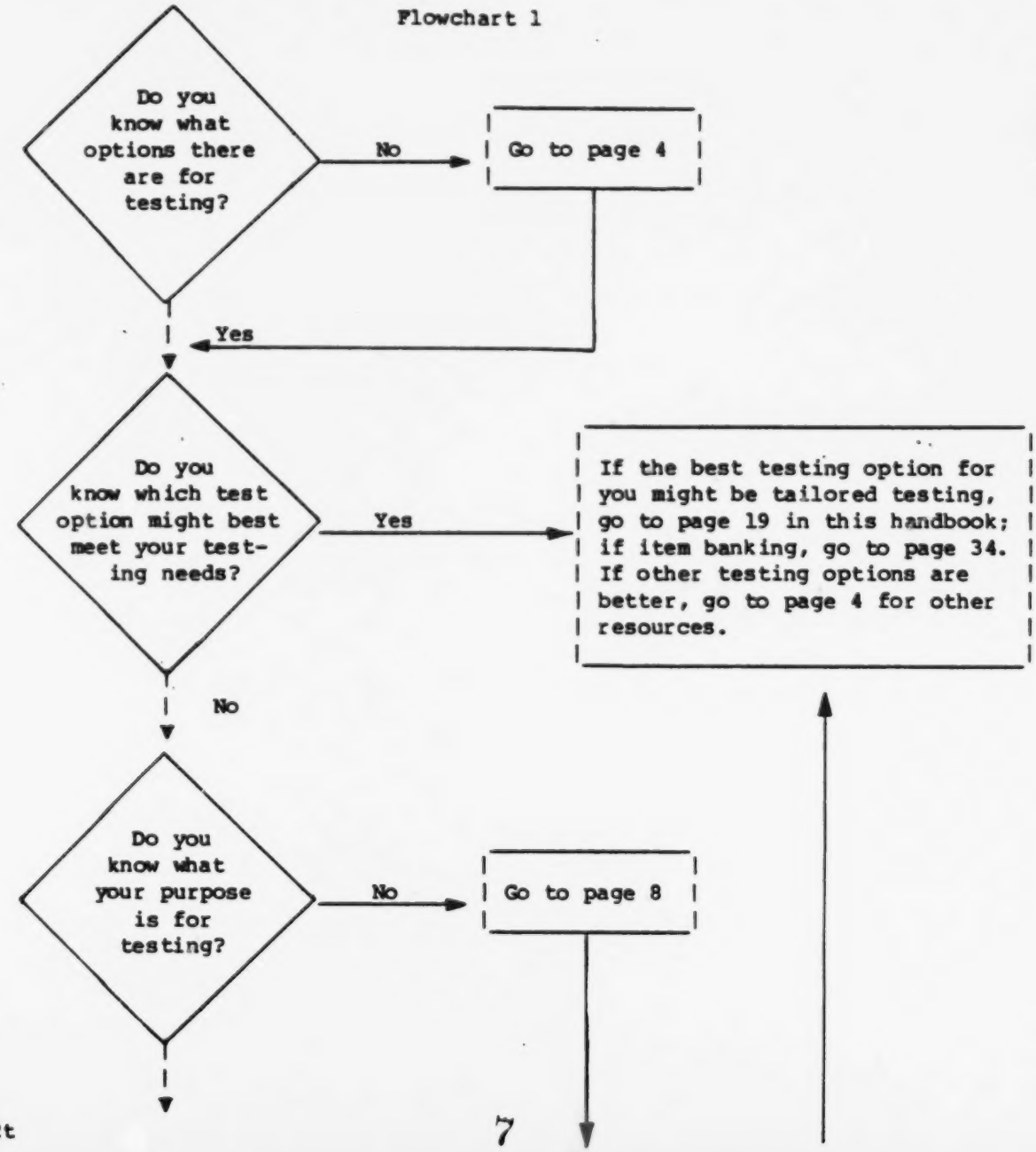
Short Form 8 -- Resources Required for Developing an Item Bank 37

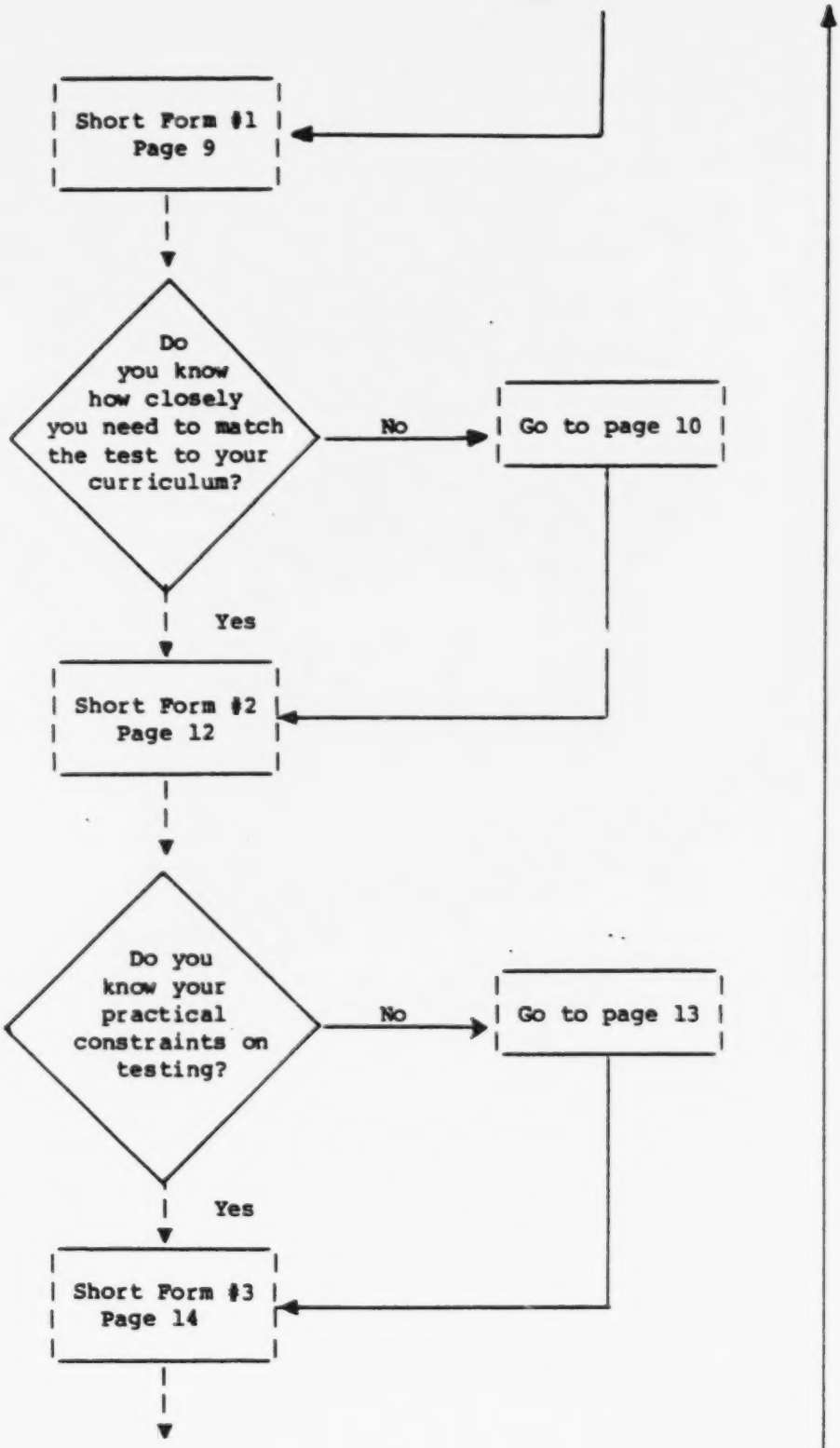
WHICH TESTING OPTIONS SHOULD BE PURSUED?

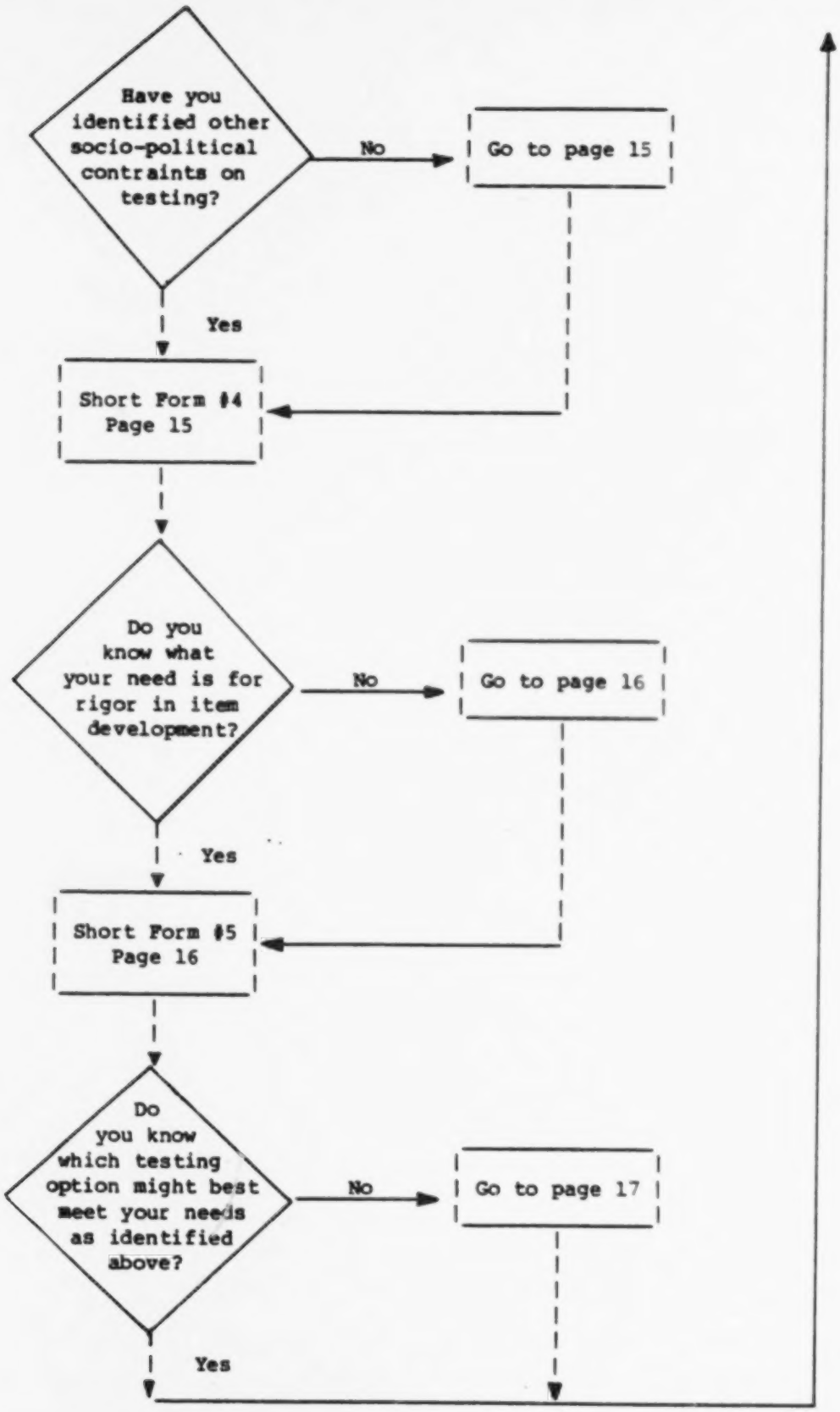
Testing options include such things as pre-packaged survey tests, curriculum embedded tests, locally developed tests, accessing someone else's item bank (customized testing) and developing one's own item bank. No one testing option will necessarily satisfy all testing needs.

Although this handbook is intended primarily to assist those who wish to pursue customized testing or item banking, some users may not know whether these options would best serve their testing needs. This section is designed to assist the reader in specifying those needs and concerns which would help in deciding whether tailored testing or item banking is a viable option, or whether testing needs could be met more practically in another fashion such as a norm-referenced standardized test.

Flowchart 1




Short Form #1
Page 9
Do you know how closely you need to match the test to your curriculum?
No
Go to page 10
Yes
Short Form #2
Page 12
Do you know your practical constraints on testing?
No
Go to page 13
Yes
Short Form #3
Page 14


Have you identified other socio-political contraints on testing?
No
Go to page 15
Yes
Short Form #4 Page 15
Do you know what your need is for rigor in item development?
No
Go to page 16
Yes
Short Form #5 Page 16
Do you know which testing option might best meet your needs as identified above?
No
Go to page 17
Yes

TESTING OPTIONS

1. Prepackaged, Survey Tests

These are generally called "standardized" tests and "norm-referenced" tests. They are the achievement tests developed by major test publishers which are intended to measure the general academic achievement of students at various grade levels. Examples of these tests are the CAT, MAT, SRA, SAT, Gates, and Nelson.

Main Uses: Student screening, survey assessment, program evaluation and guidance/counseling.

Advantages: These tests usually have rigorously developed items, measure knowledge and skills widely taught, have good norms and reporting features, and are easy to use. They usually have more than one form.

Disadvantages: Content does not always match curriculum exactly and there are a limited number of forms. Diagnostic capabilities are limited since they are given infrequently and are designed to sample from a broad range of skills rather than test a few skills in detail (although most of these tests will provide some diagnostics).

Resources: If this option is chosen, you need to screen tests for the content and features you want. Appendix E has one checklist which can be used to select a test.

2. Prepackaged, Criterion Referenced or Diagnostic Tests

These tests do not differ to a great extent from those in (1) above. The major difference is the extent to which individual skills are covered. Instead of sampling content from many areas in which students should demonstrate knowledge, as survey tests do, these tests usually focus on a smaller number of skills and cover them in more detail. These tests sometimes have norms. Examples are the Stanford Diagnostic Math Test, Prescriptive Reading Inventory, Metropolitan Achievement Tests Diagnostic Battery, or Woodcock Reading Mastery Test.

Main Uses: Student screening, infrequent student diagnosis, survey assessment, program evaluation, and guidance/counseling.

Advantages: These tests often cover skills in more detail, have rigorously developed items, and are easy to use.

Disadvantages: Sometimes the content may not correspond to your scope and sequence as well as you might like. There are usually a limited number of test forms and they can't support frequent diagnosis.

Resources: The test selection checklist in Appendix E can also be used to select diagnostic and criterion referenced tests. For such purposes, you will want a closer content match than specified in the checklist.

3. Curriculum Embedded Tests
As part of an instructional management system, such tests are intended to enable a teacher to know when a student has mastered one set of skills and is ready to move along to the next. They can be commercially or locally developed.

Main Uses: Student screening, classroom testing, student diagnosis, and mastery learning.

Advantages: These tests are supposedly directly related to the curriculum used and could satisfy need for local control. They can be used for diagnosis and mastery learning.

Disadvantages: These types of tests cannot be assumed to be of high quality, commercial systems are tied to only those curriculum materials they support, and they do not provide any norm-referenced scores.

Resources: If this is a desirable option, you can see what types of curriculum embedded tests can be obtained from the publishers of your curriculum materials, you can choose a curriculum package which has this as a feature or you can explore local development to support a local scope and sequence (see the references under number 5 below).

4. Informal Locally Developed Tests
This is the mode many teachers use. When teachers need a test to cover a particular topic, they write it.

Main Uses: Classroom testing and student diagnosis.

Advantages: The instructor has direct control over test content.

Disadvantages: Tests can be of uneven quality, development is time consuming, it is difficult to prepare parallel forms, there are no norm referenced scores, and there are no support materials for using the test scores.

Resources: Although rigorous item development is not required for this option, persons writing test items should be familiar with basic concepts for writing good items. Some important suggestions are offered in a set of workshop materials available from NWREL on writing test questions. (Also see the list of references under number 5 below.)

5. Formal Locally Developed Survey Tests
Formal self-development involves local preparation of a test (or set of tests) for some explicit and important purpose--such as survey assessment or minimum competency testing. Test development is rigorous. These become essentially like commercially available prepackaged tests in terms of use.

Main Uses: Student screening, minimum competency testing, survey level diagnosis, survey assessment, program evaluation, and guidance/counseling.

Advantages: The test content tends to match local objectives, and the tests usually are of reasonable quality.

Disadvantages: It is expensive and time consuming to develop a test in this manner if it is to be of acceptable quality. Also, there are no norm referenced scores available.

Resources: Someone who will undertake a test development should either contract with someone who has expertise in this area or obtain a book on the test development process. Examples are:

Ebel, R. Essentials of educational measurement. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1979.

Gronlund, N. Constructing achievement tests. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1982.

Gronlund, N. Measurement and evaluation in teaching. New York: Macmillan Publishing Co., 1976.

Mehrens, W. & Lehmann, I. Measurement and evaluation in education and psychology. New York, Holt Reinhart & Winston, 1984.

6. Using Another's Item Bank (Customized Testing)
Local users specify test content to another agency which pulls prewritten items. The various services and procedures available differ among agencies. The testing uses to which this procedure can be applied depend on how the item bank is set up.

Main Uses: Student screening, minimum competency testing, survey level diagnosis, mastery learning, program evaluation.

Advantages: This approach satisifies the need for local control and can result in close test-curriculum match. Items can be of better quality than those written locally, and the procedure tends to be faster and cheaper than local development. This option is often the best solution to many testing needs.

Disadvantages: This tends to be more costly than using prepackaged tests, but cheaper than developing your own. One has to use the content classification of others which may not match exactly local objectives statements. Also, because of access time, this option may not support frequent diagnosis or mastery learning unless prepackaged tests measuring each skill are developed ahead of time.

Resources: The next section of this handbook provides guidance in using an item bank developed by someone else.

7. Developing Your Own Item Bank

Sometimes it is desirable to collect and organize items locally. An item bank can be anything from a simple shoebox collection to a large computerized system accessible to many users.

Main Uses: Student screening, classroom testing, minimum competency testing, student diagnosis, mastery learning, survey assessment, program evaluation and guidance/counseling.

Advantages: Depending on the bank design, this option can support frequent and tailored testing. There can be close test-curriculum match and direct control over the testing process. Using an item bank is often faster and less expensive than writing new items each time, and the items can be improved in quality if users are willing to commit a little extra time to test analysis. Items are available from many sources (see Appendix A).

Disadvantages: Ambitious item banking projects can be expensive to set up. Using locally developed items may result in item quality problems and always require pilot testing. Care also needs to be taken in finding previously developed items because items acquired from others sometimes have not been developed carefully.

Resources: The third section of this handbook deals with issues involved in developing one's own item bank.

TESTING PURPOSES

Before deciding on any testing option, you must know your purpose(s) for testing. First, consider which groups or individuals might be using the test scores. Then ask, for whom do they need test scores? What decisions will they make using these scores? Do they have strong preferences for certain approaches, and are there some approaches they might not support? After reviewing the list of purposes, if you still have questions about possible uses for test scores in your local setting, you might consider doing a survey of potential users to determine for what they want to use test results.

Some common testing purposes are described below. These purposes primarily reflect the use of cognitive and academic test scores rather than measures of behavior or affect, although there is much overlap.

Screening

Students often have to be selected to participate in various special programs. These can include Chapter 1, special education, and gifted programs. This selection has to take place in a systematic, uniform and "fair" manner to ensure that all students have an equal opportunity of being selected, and that selection is based only on the criteria important for program participation. Test scores are often used for this purpose.

Classroom Testing

Classroom testing refers to those in-class tests used to measure student progress, often for grading. Such tests include quizzes, midterms and final exams, as well as drill-and-practice sheets.

Certification Testing

Certification testing (sometimes called minimum competency testing) is used to make decisions about student promotions and graduation, or to decide which skills students have mastered to provide students with remediation. The purpose is to see whether the student demonstrates some acceptable skill level required to function in some setting such as the next grade level or the adult world.

Student Diagnosis

Teachers often want detailed information on which skills and subskills students have mastered. Such information helps teachers in developing instructional programs that genuinely strengthen student performance.

Mastery Learning

Some educational programs are set up so that students must demonstrate mastery of one skill before they can proceed to the next. A series of small tests embedded in the curriculum provides this information. Also, general skills tests at the beginning of the school year can be used to place students in the curriculum.

Survey Assessment

Survey assessment is used to make overall statements about how well a defined group of students is doing with regard to specified skills. Such broad-based assessment often functions much like a regular checkup at the doctor's office--there's no reason to believe that anything is wrong, but if there is, you want to have an early warning. In general, survey assessment samples the performance of students on a variety of skills and objectives which are of local or national importance. Performance is checked against "typical" (i.e., norm referenced) performance or "ideal" (i.e. criterion referenced) performance so that the community can decide whether the educational program is on track.

Program Evaluation

Testing is often used to assess the adequacy of educational programs, methods and materials. It is also used to make judgements about the effects of particular program components.

Guidance and Counseling

Test scores are often used to assist professionals in making clinical appraisals of students, as well as suggesting to students courses of study and areas of interest. Although affective and behavioral measures are often used for those purposes, tests of academics are also used.

Implications for Item Banking

In general, item banking is more useful for mastery learning and diagnosis than the other testing purposes.

Short Form 1

The purposes for testing in my local situation are (rank order all that apply):

____ Student screening
____ Classroom testing
____ Minimum competency testing
____ Student diagnosis
____ Mastery learning
____ Survey assessment
____ Program evaluation
____ Guidance and counseling
____ Other _____

TEST-CURRICULUM MATCH

As we've already noted, you must have the purpose for testing clearly in mind when deciding on testing options. Many considerations in addition to purpose can help you narrow the range of possible choices further. One of the most important is the need for test content to match what is taught. Major standardized tests differ in their specific content because they are based on broad samples of information at each grade level. Thus, the extent to which they are likely to match the specific content of any single curriculum's scope and sequence is relatively unpredictable. The extent to which this is desirable or undesirable depends on how scores are to be used.

The central issue here is one of inference. We use test scores to infer how well students have mastered the domain of content from which the items are sampled. We can't test everything. A test score represents performance on only a sample of all the possible test questions which could be written to assess a skill. The question is whether the sample of questions on the test will adequately support inferences we wish to make. Inferences to specific skills, such as "knowledge of beginning vowel sounds" requires fairly specific test questions. Inferences to more general skills such as "knowledge of phonics" requires a broader sampling of skills. Finally, a general inference such as "does our reading program promote comprehension" requires still a broader sampling of skills. For more discussion on test-curriculum match see the Fall, 1984 issue of Educational Measurement, Issues and Practices.

The following purposes usually call for a relatively close test-curriculum match:

1. Classroom testing. If the teacher is testing for grading purposes, the question of inference is whether students learned what was taught; therefore, the teacher will want to sample information directly from what was taught. Specific content should be represented on the test in direct proportion to its importance in the course.

2. Minimum competency testing. The definition of skills as "minimum" implies that all students should master them. There may be some minimums taught but not tested, but all tested skills should be considered essential. Such important educational decisions give little leeway for "extra" information on the test. In addition, the information on the test must validly sample the essential skills.

3. Student diagnosis. Diagnosis is most effective when related to a particular scope and sequence. If there is no locally defined scope and sequence, it might be most helpful to use a packaged test covering skills typical at a particular grade level. Also, in order to make inferences about needs on individual skills, the skill should be measured by at least three items, and preferably more.

4. Mastery learning. Tests designed to measure whether particular skills have been mastered before a student proceeds to the next level need to be directly tied to scope and sequence, and to cover each skill thoroughly.

The following testing purposes generally do not require such a close match to the local curriculum. However, local circumstances may alter this judgment.

1. Student screening. Student screening often requires only that the best (or slowest) students be identified. Any method of rank ordering students accurately will work. It is not so crucial that test content match local curriculum here since students will probably rank order the same regardless. An exception might occur where students are screened into a program based on some particular skills which all need to be measured. An example is the case when students are screened for remediation of specific basic skills.

2. Survey assessment. Often the question in survey assessment really is "How are we doing compared to others in the country?" If this your question, only a norm referenced test will give you the answer. Such tests are not unrelated to specific programs, but often contain some items covering information not taught, or leave out some items on information which is taught. This can actually be desirable if the question asked is "Does our scope and sequence teach skills generally considered important?" If a program is so "unique" that a standardized, norm referenced test does not measure progress at all, perhaps the program should be redesigned. Also, many local people want to know strong and weak areas--even if the areas are not stressed in their curriculum, e.g., "We don't stress these skills, but we are 'curious' how we are doing on them."

   At other times the survey question is "How are we doing on the skills which we have designated as most important?" In this case, the match between instruction and test content might be more crucial.

3. Program evaluation. Again, the evaluation question being asked is critical. Suppose you're asking, "Does this program component teach skills a, b, and c better than another program component?" In that case, it is critical that the test cover skills a, b and c. More general questions such as "How well does our program teach reading?" may allow for a more general measure. The evaluation design may also influence which type of test is most desirable. For example, when using the norm referenced model in Chapter 1 evaluation, you need a test with norms.

Technical reasons aside, sociopolitical pressures may demand a close match between test and curriculum. For example, the local climate might result in any mismatch receiving an inordinate amount of attention, and detract from constructive use of the results.

Implications for Item Banking

In general, with increasing concern about testing specific content, there is increasing need to either carefully select a prepackaged test, or tailor your own test.

Short Form 2

My need to match test content to what is taught is:

____ Low (we only need general measures of content in the same general areas)

____ Medium (at least 75% of the items on the test should measure content we cover and at least 75% of our major objectives should be covered by the test)

____ High (all test items should directly relate to a specifically designated skill or objective, and all skills and objectives should be represented on the test in proportion to their locally judged importance)

PRACTICALITIES

In addition to testing purpose and need for test-curriculum match, practical constraints on testing can influence what testing option will be used. Practical concerns relate to required testing frequency, the need for individually tailored tests (i.e., different tests for different students), the need for multiple, equivalent forms (i.e., multiple items covering the same content), outside constraints on testing (such as state and federal requirements), and the need for particular types of test scores (such as percentiles).

Testing Frequency

Testing purposes which generally require less frequent testing (one to two times a year) are screening, minimum competency testing, survey assessment and program evaluation. These generalizations don't always hold. For example, sometimes diagnosis occurs infrequently, as when tests are given at the beginning of a school year to place students into a scope and sequence. Sometimes minimum competency testing occurs more frequently, as when students are invited to demonstrate individual competencies at any time during the school year.

Individually Tailored Tests

Generally speaking, test content is individualized only in diagnosis and mastery testing. For other purposes, students will generally receive the same test items. (Although there have been recent developments on computer-administered testing which might result in more individually tailored tests for all uses.)

Multiple, Equivalent Forms

Parallel test forms are necessary when students must be retested on the same general content without responding to exactly the same items each time. Such is the case for minimum competency testing and mastery learning. In some cases, since progress is being measured through the curriculum, students might need to be tested repeatedly until they demonstrate skill acquisition. Parallel forms ensure both security and accuracy.

Test Scores

Sometimes, certain types of test scores are required to meet a specific need. If the testing questions deal with norm group comparisons, percentiles are called for. If the testing questions deal with mastery, test scores must be translated into mastery statements.

Resources

Sometimes the range of testing options is constrained by limited resources: money, time, expertise and equipment. Low resources generally imply prepackaged tests which are easy to give, score and interpret; increasing resources widen the range of options. While resources do not determine which testing option is "best," they often place a ceiling on what can be done. For the time being we will concentrate on the "ideal" testing options--recognizing that what is ideal or desirable may <u>not</u> be possible because of resources.

Implications for Item Banking

In general, item banking becomes more feasible with increased need for frequent testing, individually tailored tests and multiple equivalent forms.

Short Form 3

My practical considerations are:

1. Testing frequency
   ____ Low (one to two times a year)
   ____ Medium (three to four times a year)
   ____ High (five or more times a year)

2. Need for individually tailored tests
   ____ Low (never or hardly ever)
   ____ Medium (occasionally)
   ____ High (frequently)

3. Need for multiple equivalent forms
   ____ Low (can use the same test each time)
   ____ Medium (would like two to three forms to rotate)
   ____ High (must have a different set of items each time)

4. We need the following types of test scores:
   ____ Percentiles/NCEs/Stanines/Grade equivalents
   ____ Number right/percent right
   ____ Mastery statements

SOCIOPOLITICAL CONSIDERATIONS

The local testing context can influence test selection via several factors:

1. Local opposition to "standardized" tests. This usually comes down to a feeling that "those tests don't measure what we teach."

2. Concern that tests will dictate the curriculum.

3. Concern that existing tests cannot measure many, if not most, important educational outcomes.

4. Suspicion that testing is too secret, that items are not revealed to the public so their content and quality can be judged openly.

These concerns usually boil down to a general "need for local control" over test content, test items, and the testing process in general.

Implications for Item Banks

In general, item banking is more feasible in those environments which desire local control over the testing process.

Short Form 4

Our need for local control over the testing process is:

____ Low

____ Medium

____ High

NEED FOR RIGOROUS ITEM DEVELOPMENT

Rigorous item development implies that care has been taken to ensure that items measure what is intended, that no extraneous features--such as wording or response length--influence a student's choice independent of knowledge, and that the items perform in a desirable manner in actual use. One can have reasonable assurance that test items which have undergone sufficient scrutiny and pilot testing are of good quality.

This process, however, is time consuming and costly; moreover, it requires a good deal of expertise. Therefore, one needs to identify the situations in which such care is essential, and those in which such care is less important. In general, when test scores are going to be used to make important, lasting decisions about students or programs, test items need to be very carefully prepared. Such decisions might involve, for instance, screening of students for certain programs, minimum competency testing, survey assessment and program evaluation. Decisions involving diagnosis and classroom grading--that is, those which are reversible or less critical--may not require such rigorous scrutiny of items.

Implications for Item Banking

In general, need for rigor in items either implies using a good prepackaged test or use of carefully scrutinized items when doing item banking.

Short Form 5

My need for rigorous item development is:

____ Low

____ Medium

____ High

WHAT TESTING OPTIONS MIGHT BEST MEET MY NEEDS?

Based on your responses to the questions in the Short Forms, you can use the table on the following page to find out which testing options might match your needs. (Testing options are described on pages 4-7.) Check your responses for each area against the responses typical for that option. Then choose the one or two closest matches to pursue.

For example, let's assume then you do not have a lot of resources, you want to do student selection and survey testing (a and f on the chart), you want as close a match to your curriculum as possible, you will test only twice a year, you do not need multiple equivalent forms or individually tailored tests, you would like percentiles, there is not a major feeling about local control, and you want rigorous item development. Your pattern of responses is: M, L, N, L, H, L, M-H and testing purposes a and f. Your most likely choice would be a prepackaged survey test--or Option #1.

<u>Caution</u>: We do not mean this process to be totally prescriptive, but rather to provide a means of judging "best bets." Many local considerations not mentioned in the previous sections could <u>influence the final decision</u>.

Table 1

Deciding on a Testing Option

Consideration	Your Rating	Test Options*						
		Prepkg Survey Tests with Norms	Prepkg CRT or Diag. Tests	Curric. Embedded Tests	Informal Locally Develop. Tests	Formal Locally Develop Tests	Custom-ized Tests	Item Banking
Need for Test-Curr. Match		M	M-H	H	H	H	H	H
Frequency of Testing		L	L	M-H	L-H	L	L-M	L-H
Need for Indiv. Tailored Tests		L	L	L	L-M	L	L-M	L-H
Need for Mult. Equiv. Forms		L-M	L-M	L	L	L-M	L-H	L-H
Need for Percentiles		L-H	L	L	L	L	L***	L***
Need for Local Control		L	L	L-M	L-H	L-H	L-H	L-H
Need for Rigorous Item Develop.		L-H	L-H	L	L	H	L-H	L-H
Testing Purpose(s)**		a,d, f,g h	a,b, d,e, g,h	a,b d,e	b,d	a,c, d,e, f,g,h	a,c, d,e, f,g,h	a,b, c,d,e, f,g,h

*See pages 4-7 for descriptions of test options.

**Testing purposes listed are those that might match each test option. Codes are:

a = Student screening
b = Classroom testing
c = Minimum competency testing
d = Student diagnosis
e = Mastery learning
f = Survey assessment
g = Program evaluation
h = Guidance and counseling

*** This rating refers to the general state of affairs in most item banks. There are many ways to estimate percentiles in <u>some</u> item banks if all the items are calibrated.

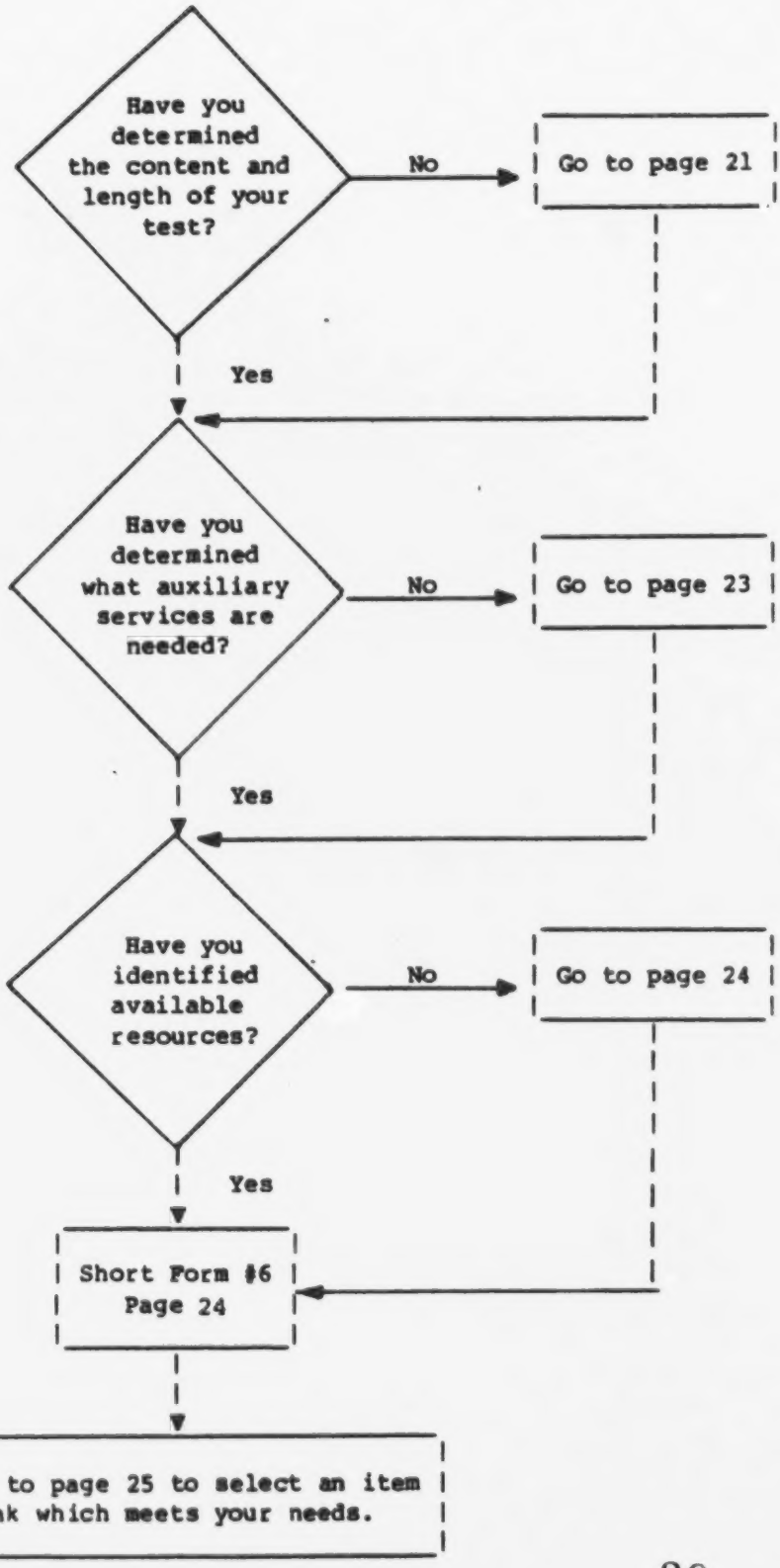
USING ITEM BANKS MAINTAINED BY OTHERS
(Customized testing)

This section is intended to assist you in selecting and using an item bank to have tests made for you. Item banks which provide this service often call it "customized" or "tailored" testing. Many groups, including test publishers, school districts, consortia of districts, and state departments of education, provide a customized testing service. In general, the process is that you use the bank's item classification system to specify which skills are to be tested and the number of items for each. Since you are a user of a bank developed and maintained by someone else, you have no direct control over which items are in the bank or how they are classified. You simply access what is there using whatever content classification system they have developed.

In addition, various item banks will provide more than just items. Sometimes they offer test scoring, reporting results, cross-referencing of skills to instructional materials and other things. In order to choose an item bank to use, you need to therefore consider those "auxiliary" functions as well as content.

This section of the handbook will assist you in defining your testing needs if you use another's item bank. Examples of customized testing using various item banks are on yellow pages 32-33.

Flowchart 2


Have you determined the content and length of your test?
No
Go to page 21
Yes
Have you determined what auxiliary services are needed?
No
Go to page 23
Yes
Have you identified available resources?
No
Go to page 24
Yes
Short Form #6
Page 24
Go to page 25 to select an item bank which meets your needs.

WRITING TEST SPECIFICATIONS

The primary benefit in developing your own tests from an item bank maintained by others is your ability to test the content important to you in a way that meets your needs without having the cost of developing and maintaining the item bank yourself. But, it also has a drawback in that rarely will you find an item bank which stores its items by skill objectives stated exactly like yours. Therefore, you need to develop specifications for your test(s) so that you get the items you want.

Test specifications are the blueprint for your test. To pull those items from someone else's item bank that meet your testing needs you must know what content the items need to cover and how many items of each type to request. The content requested needs to relate directly to your curriculum and instructional objectives. The number of items depends on the relative importance of each objective you want to test and the total length of the final test. Thus, first you specify the skills to be tested in terms of your own local objectives, and then you translate your objective statements into the content classification statements used by the item bank so they can find the items you want.

Yellow pages 27-31 in this section provide three examples of test specifications. The information below provides some hints on how to develop test specifications.

Sources for Determining Test Content

- Scope and sequence documents
- Competencies for each grade
- Textbooks and other materials used
- Survey of potential test users--who may need various kinds of information
- Content classification schemes of various item banks

Other Things to Consider in Specifications

When developing test specifications, you may want to list other special characteristics that items should have. Possibilities are:

- Response format--multiple-choice, matching, true/false, short answer, essay.
- Item difficulty--the number of students in a given category who have, in previous test administrations, answered the questions correctly.
- Cognitive level--e.g., recall of knowledge, inference or application of knowledge, as in Bloom's taxonomy.
- Specific topics--e.g., you might want reading passages of particular types such as an essay, story or recipe.

These and other special characteristics of each item should be listed.

Hints

1. Be specific. Instead of saying that first graders will have test items on "phonics," say "beginning and ending consonant sounds." Specificity will not only help users decide what should be on the test, but will also help in selecting items from the item bank because the content descriptions for the items on the bank you are accessing may not match precisely with your local scope and sequence. You must usually match your content descriptions with the bank's.

2. Be careful. The care with which the test specifications are outlined should be directly proportional to the importance of the educational decisions to be made. If, for example, the test will be used to certify students for graduation, you must be very sure that the test content accurately reflects what people think is important, and that students have had an opportunity to learn the content assessed.

3. Use common sense. It is not always more efficient to select from existing items. Sometimes it is easier to write items than to obtain them through an item bank. This usually depends on testing purpose and content. For example, many math computation items are very easy to write. If, however, the test will be used for important educational decisions you might want to use existing items anyway if they have undergone more rigorous development and trial testing.

AUXILIARY SERVICES

Publishers offering customized tests using their own item banks vary in the assistance and services they provide. Of course, the more services desired, the greater the cost. But, you might as well have a "wish list," it can always be pared down later.

Item Typing and Test Printing

Some publishers offering customized testing will only provide items. You need to do the work of formatting and typing the items, adding instructions and sample items, and printing the final tests. Others will completely format and print the tests once you have selected the items you want. You should find out how the items will come to you.

Item Review

Find out how the final items for the test will be selected. Some publishers offering customized testing will let you review the items they select from their bank to approve the selection or delete items and request others. Others will send you the items with no provision for review.

Scoring

Be sure that your item source at least sends you an answer key. In addition you can sometimes arrange for test scoring. This service is commonly offered by commercial test publishers. You might be especially interested in this option if you need scores broken down in various ways--for diagnosis or mastery learning, for instance.

Reporting

If you want to have the tests scored for you, find out what reports are available. You will probably want at least student level results and class summaries. In addition you might want special reports for various subgroups of students, using different types of test scores, etc. Norm referenced comparisons will often not be available from item banks. Many, however, provide other interesting performance measures such as placement of students along a continuum of skill levels (ask if item calibrations, e.g. Rasch, are available).

Referencing Curriculum Materials to Test Results

Some item banks have cross-referenced their test items to curriculum materials in which the items' concepts are covered. This might be of special interest to those who are developing tests to assist in diagnosing student skills.

RESOURCES

In planning your testing system, you need to know what resources you have available. This will determine in great measure what you will do yourself and what you will ask of the item bank.

List your resources in the following areas:

Short Form 6

	Your Resources
MONEY (costs could include buying items, formatting printing, scoring tests and reporting results)	
TIME (time could be spent assessing testing needs developing test specifications, finding out about available banks which tailor tests, reviewing items, and formatting the test)	
EQUIPMENT (you might need a word processor or a scanner for answer sheets)	
EXPERTISE (you might need personnel who can assess item quality, interpret item statistics, lay out and paste up tests, write sample items and instructions, and develop a test administration manual)	
COMMITMENT (you will need the backing of those who will use the tests)	

SELECTING AN ITEM BANK

Criteria for selecting an item bank are listed down the side of Form 7. List the item banks you are considering to provide you with customized testing along the top. Then either (1) rate all the item banks on a scale of one to five on each criterion or (2) list information about each item bank in the space provided. The various criteria are briefly described on the next two pages. A list of item banks appears in Appendix A. This information is abstracted from a survey done by NWREL in 1984 entitled A Guide to Item Banking In Education (Second Edition). Appendix A will assist you in determining which item banks provide customized testing, what services are available from these banks, and what content areas each covers. (Note: This list of item banks is not necessarily exhaustive, but represents those who returned surveys. You might also try other test publishers. In addition, the summary of survey results in the back may not provide you with the final information you need to make a selection. You need to narrow potential choices down to two or three likely candidates and then call them for final details. Finally, you will note that many of the agencies are public, not commercial, test publishers. While they are willing to share and cooperate, it will be best to have a cooperative approach which considers the demands on their time and resources.)

Short Form 7

	Item Bank			
Selection Criteria				
1. There are enough items covering desired topics and grade levels.*				
2. Desired auxiliary services are available.				
3. Items are categorized in a desirable manner.				
4. Tests are developed from the bank in a desirable manner.				
5. Items in the bank have had the needed quality checks.				
6. Cost is appropriate, given available resources.				

*If the minimum number of items for your subject and grade is not available, the item bank is not useful--don't rate.

Explanation of Criteria

1. Number of items in desired topics and grades.
   Your test specifications will describe test content and the number of items required for each area. Appendix A lists various item banks and the content covered by their items. The more items in a bank, the greater the likelihood you will get a good match to your desired content. If the minimum number of items for your subject and grade is not available, the item bank is not useful for you--don't consider it further.

2. Availability of desired auxiliary services.
   Appendix A also provides a brief view of the various auxiliary services available to users from the item banks that answered our survey. Auxiliary services include test printing, scoring and reporting.

3. Items are categorized in a desirable manner.
   You may want to select items on the basis of information other than content and grade. Such characteristics should be included in your test specifications. Make sure that the item bank allows selection of items on all criteria you want. Appendix A indicates some item selection possibilities for item banks which returned our survey.

4. Tests are developed from the bank in a desirable manner.
   You need to match turnaround time for obtaining the test to your testing schedule. Long turnaround times will not, for example, support frequent testing. Another consideration is the procedure for review and selection of items. (See page 23 for a discussion of procedures.) A final consideration is ease of access--is it easy to use the bank?

5. Items in the bank had the needed quality checks.
   Procedures for reviewing and entering items into banks differ. Some banks include every item they get with little or no screening and/or field testing. Other banks go through an elaborate (and costly) review process. The level of item review you need depends on the importance of the test. For important educational decisions, such as promotion and placement in grades or special programs, the items should be of the highest quality. You might consider any combination of the following as critical for your items: (a) pilot tested, (b) reviewed for sexual, ethnic or cultural bias, (c) technical editing, and (d) review for content and/or grade level match. Some of this information is available on the survey summary in Appendix A.

6. Cost
   Once you have developed your "wish list," you can compare availability and cost to revise the list to fit your budget. The goal is to obtain all essential services and as many desirable services as possible within cost, quality and efficiency constraints.

SAMPLE TEST SPECIFICATIONS

Three examples of test specifications are on pages 27-31. Common features are that they provide information on:

1. The skill(s) to be tested.
2. Sample items as models for the ones to be selected.
3. The number of items to be selected to measure each skill.

Other information could be added or cross-referenced as needed. For example, in the specifications on page 30 there are indications of the level of cognitive processing to be tested by each item.

Test Specification - Sample A

AREA:	Mathematics Computation	GRADE LEVEL: 4
TOPIC:	Arithmetic Word Problems	
SUBTOPIC:	Addition of Whole Numbers	

OBJECTIVE:	Given a mathematical word problem involving addition of whole numbers not greater than four digits, the student will select the correct answer.
WEIGHTING:	30% (approximately 15 items)
ITEM TYPE(S):	Multiple-choice

EXAMPLES OF ITEMS

ITEM 1: John has 312 stamps in his collection, Greg has 224, Pete has 101 and Bob has 252. How many stamps do the boys have altogether?

a. 798
b. 789
c. 879
*d. 889

ITEM 2: Ed's book has 144 pictures, Susan's book has 21 pictures, Jim's book has 33 pictures. How many pictures are there in the three books?

a. 54
b. 98
c. 177
*d. 198

Test Specification - Sample B
Description of Situations/Displays of Items

Description of Situation/Displays of Item Types	Competency to[a] be Measured
1. Line Graph--unemployment in US vs. St. Louis (Form 1, p. 12[b])	
a. Describe a trend	1
b. Read a value	1
c. Make a comparison (Form 2 #86)	1
2. Bar Graph--percent women in labor force--State vs. a company (Form 2, p. 13)	
a. Make a comparison (Form 2 #61)	1
b. Read a value	1
c. Make an inference (e.g., What conclusion could be supported by this display?)	
3. Pie Graph	
a. Make a comparison	1
b. Read a value	1
c. Add percentages	5
4. Map (island road map--with symbol legend)	
a. Read distances (distances between points will be written in--no calculation required)	1
b. Find a location using the sumbol legend	7
c. Find the airport or hospital using the symbol legend (e.g., Near which city is the airport?)	7
d. The best way to get from point A to point B	1
5. Job ad (reading level--grade 6)	
a. Find a piece of information	1
b. Identify skills needed for this job	12
c. Differentiate fact vs. opinion	9
d. Identify needed interests, locations, goals, etc., required for this job	12
e. Write a letter in response to this ad--content and stype	3

[a]This is a cross-reference to whatever set of competencies are being locally used.

[b]This is a cross-reference to sample items or displays which might provide a model when developing the test.

Test Specifications - Sample C

Purpose

This document is intended to guide the item writers in finding/writing items for the Model Life Skills Tests. The test specifications are intended to reflect the suggestions, priorities and emphases of the advisory panel which met March 5 and 6 to discuss the content of the tests.

Overall Test Development Considerations

The advisory panel provided some overall approaches/philosophies which will guide the test development process. These include:

1. These tests should not be the same as a regular achievement test. They should not reflect skills in an "academic" manner but should reflect the application of these skills in adult life. For example, the math items should reflect problem solving situations from everyday life. These often require much estimation, rounding, and several steps. As another example, the passages used for the reading items should be taken from everyday materials such as the newspaper, written information on traffic tickets, advertisements, guarantees and instructions.

2. All items should be "in context." That is, all items will be related to a real life situation. There will be no lists of math computation problems, and no lists of vocabulary items.

3. This is not a "minimum competency" test. Therefore, the items should reflect the same range of difficulty that persons are likely to encounter in real life.

4. The material should be "regionalized." For example, articles should be taken from Washington newspapers, Washington forms should be used, and place names should look like they are from Washington.

Sample Specifications for the Reading Test

Domain 1: Reading, understanding and using written material from everyday life. (40 items)

Possible Situations/Displays (at least one display on each major category):

a. Instructions--how to put something together, how to do something, prescriptions, recipes, how to go somewhere

b. Warnings--poisonous household goods, street signs, medicine labels, safety signs

c. Information/instructional material--nutrition, reference books, pamphlets, driver's manual, job announcement, clothing care labels, microcomputer manual/tutorial

d. Leisure materials--menu, magazine articles, newspaper articles, movie/television listings, correspondence

e. Legal documents--wills, insurance, public notices, guarantees, traffic ticket

f. Work related--vouchers, requisitions, work orders, bills, correspondence, employer handbook, safety manual, reports

g. Persuasive material--advertisements, speeches, editorials

h. Forms

Item Types

a. Finding information/retrieving facts or details--12 items

b. Sequence of events--4 items

c. Identify fact versus opinion--4 items

d. Interpreting the reliability of various sources--4 items

e. Making inferences (comparing/causation/predicting outcomes/noting inconsistencies)--8 items

f. Identifying main points and subsidiary ideas--5 items

g. Identifying writer's purpose in a passage written to inform or persuade--3 items

Domain 2: Understanding the meanings of words used in common reading situations. (20 items)

Possible Situations/Displays

All vocabulary items would occur in the context of the reading passages chosen for Domain 1.

Item Types:

a. Choosing the best definition of a word--5 items

b. Determining the meaning of a word from context-5 items

c. Identify antonyms--5 items

d. Prefixes and suffixes--5 items

Considerations in Developing the Reading Test:

1. The general readability of passages will not be controlled. Passages will be selected from materials which graduates will encounter in everyday life. These will represent a range of reading difficulties.

2. Vocabulary items will be in the context of the passages used for the reading comprehension items. Thus, each passage will be followed by both comprehension and vocabulary items. The words in passages chosen for vocabulary items should not be above grade 10 in difficulty unless the purpose of the item is to deduce the meaning of the work from context. A good idea is to underline the words in the passage which will be subsequently used in vocabulary items.

3. There will be no short stem, out-of-context type items.

4. The passages should differ in length, but none should probably be more than 200 words long.

CUSTOMIZED TESTING EXAMPLES

The following two pages give examples of how customized testing proceeds in three different item banks. The purpose is to show the general steps involved and the possible differences in processes between item banks.

Example 1--Customized Testing From a Test Publisher

> "From lists of objectives, the district can select the performance objectives to be tested and specify the number of test questions to measure each objective. (The test publisher) will then prepare customized test booklets from its extensive bank of multiple-choice test items to match the selected objectives. After the tests have been administered, they will be scored and the results reported in a variety of criterion-referenced formats."

In this example, users must match their own test specifications to the publisher's item categorization. Items can be selected by objective, difficulty and/or cognitive level. It is not clear whether there are provisions to review items before they are compiled into the final test booklet. The publisher formats and prints the test. There are options for scoring, a variety of reports available, and assistance in selecting objectives. There are over 1,100 reading items, 2,000 math items, 800 language arts, and 300 other items available for grades K-12.

Example 2--Customized Testing From a State Department of Education

The purpose of this state item bank is to assist districts in the development of state-mandated competency tests used to evaluate student progress at specific times, and to develop remediation plans as needed. Flexibility is allowed in test content and, to some degree, in the grades at which testing occurs. District use of the item bank is optional.

In this scheme the district develops its curriculum and competencies list, then matches the competencies to categories within the item bank. Items in the topics requested are then downloaded from the main data bank into a microcomputer. The district makes an appointment to examine the items on a microcomputer at the state department, and marks desired items. The state then provides the district a copy of the items by competency. Once the district has reviewed items and made final selections, the state pulls items, formats the test, and sends the final test to the district. There are provisions for reexamining competencies and rematching needs to the item bank's categorization scheme. The state produces camera ready copies of the test, complete with administration instructions and answer keys. The district can add its own items to the test but the state does not print them nor enter them into its item bank. The district is responsible for its own test administration, scoring and reporting.

It takes about 1-2 weeks to prepare requested topics for viewing by the district. Once the initial list of questions is developed, it takes about one week to send draft items to the district. When the district has approved the items it takes about 1-2 weeks to get the camera ready copies. The service is free to districts in the state.

Example 3--Customized Testing Using a County Item Bank

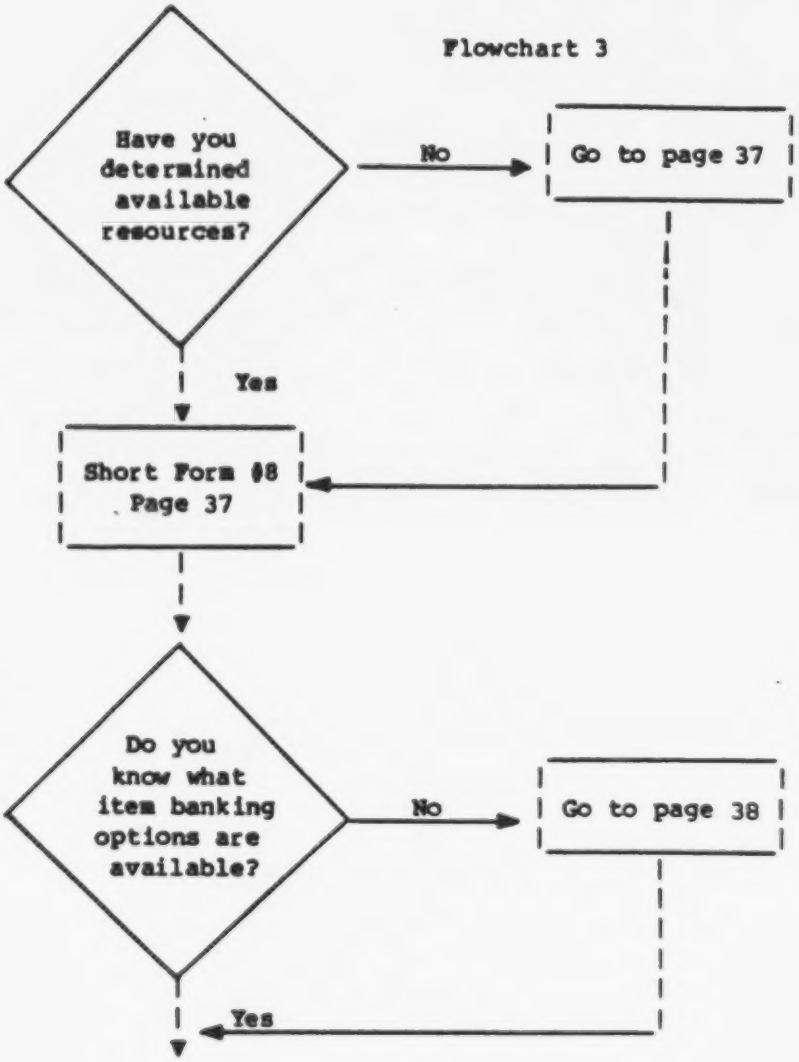
This county item bank is intended for instructional management and mastery learning applications by schools and individual teachers. There are extensive basic skills scope and sequences to which items are referenced. There are three types of tests that teachers can obtain. The first type is a prepackaged, global grade level test which generally places each student within the scope and sequence. The second prepackaged type is designed to measure specific subskills relevant to any areas in which a student might be weak. Under the third option, teachers can obtain tests customized to specified skill areas. Items can be pulled by direct access to the item bank through a computer. The system allows teachers to add items to any test. Tests can be scored by the system and a variety of individual student and classroom summary mastery reports generated. The system, implemented on a large computer, took four years to develop.

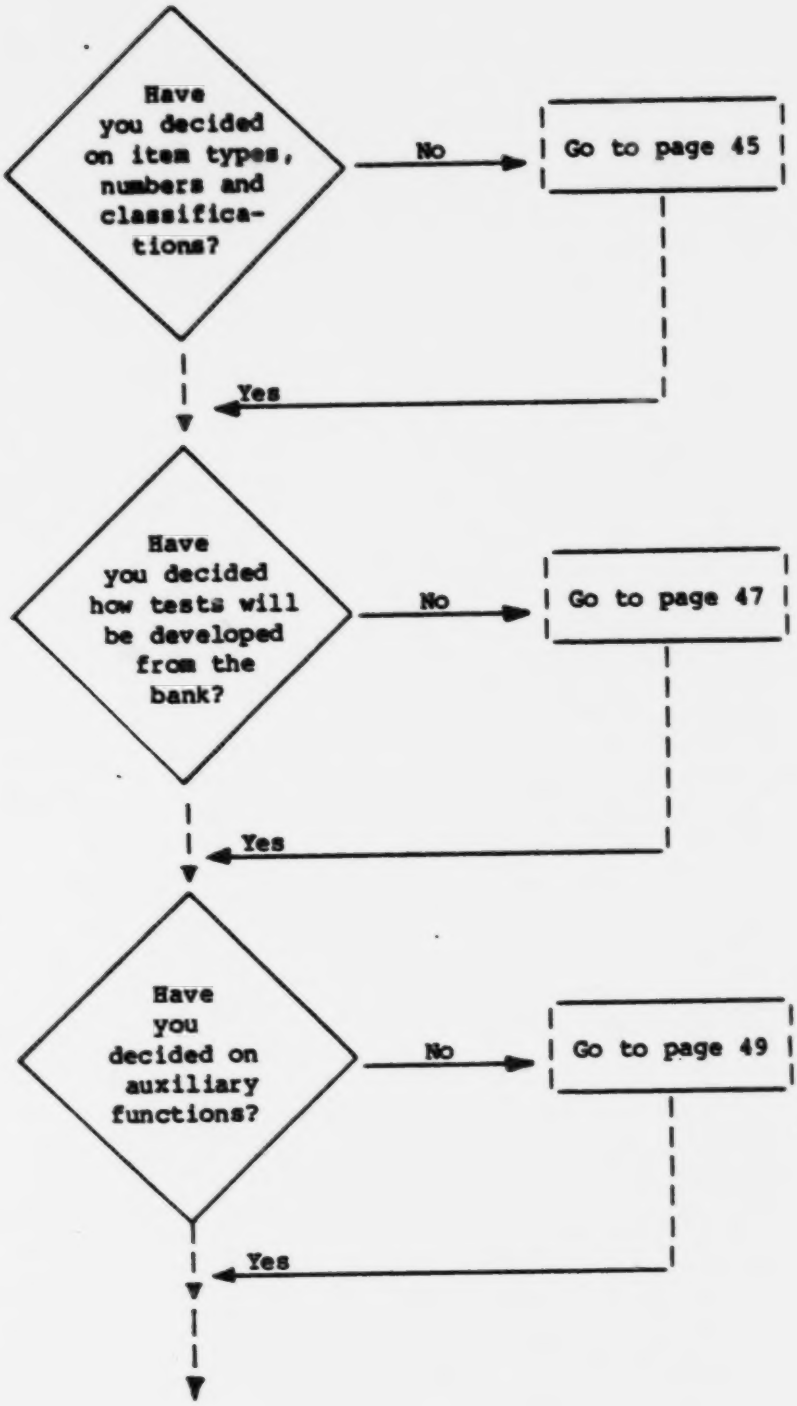
This item bank attempts to address the teachers' needs for diagnosis and mastery learning tests that provide for quick turnaround times. Turnaround for tests developed centrally is 3-8 weeks. Test scoring and reporting is generally 2-4 weeks.

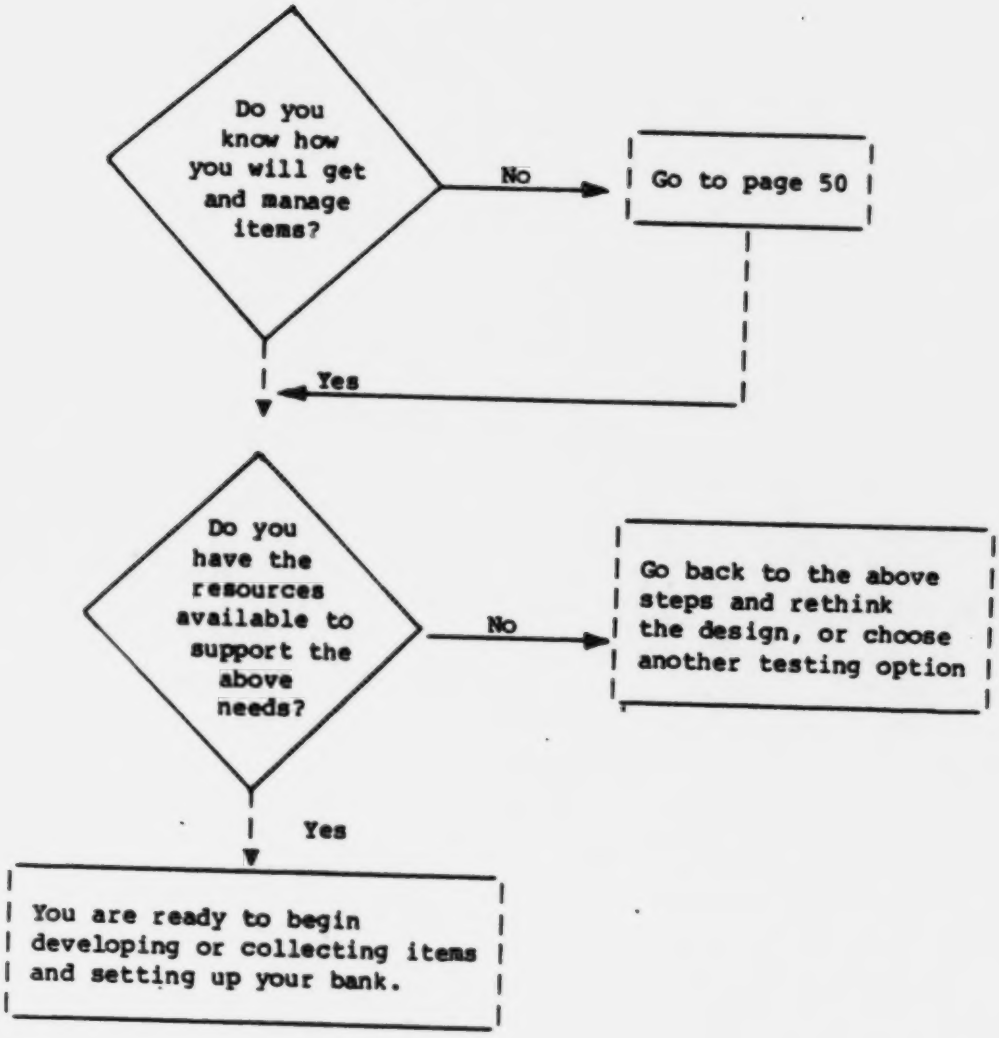
DEVELOPING YOUR OWN ITEM BANK

This section of the handbook covers the major decisions, considerations and options involved when developing your own item bank. It is intended to assist those who find that developing and maintaining their own item bank is a viable option for meeting their testing needs. This section can also be used to help readers decide whether or not item banking is their best bet.

Flowchart 3




Have you decided on item types, numbers and classifications?
No
Go to page 45
Yes
Have you decided how tests will be developed from the bank?
No
Go to page 47
Yes
Have you decided on auxiliary functions?
No
Go to page 49
Yes


Do you know how you will get and manage items?
No
Go to page 50
Yes
Do you have the resources available to support the above needs?
No
Go back to the above steps and rethink the design, or choose another testing option
Yes
You are ready to begin developing or collecting items and setting up your bank.

0712t

WHAT RESOURCES ARE AVAILABLE?

Item banking can often require substantial resources. Major up-front item acquisition and classification (not to mention software and hardware acquisition) can be costly. Whether this type of up-front development is required can depend on your needs. (For example, large systems with multiple users can require that the system needs to be completely developed before use is begun.) In any case, you need to know what resources are available to you in order to decide how you might set up a bank given local constraints.

List your resources in the following areas:

Short Form 8

	Your Resources
1. Money available	
2. Staff time available	
3. Equipment (computers, word processors, Xerox, etc.) available	
4. Expertise in the areas of computers and test development	
5. Staff and administrative commitment	

ITEM BANKING TYPES

Item banks come in every size and shape. Six types are listed below and described on pages 40-44. Specific examples are provided on yellow pages 52-56. These types reflect a continum of complexity--from simple manual systems to sophisticated computer assisted systems. The exact configuration of your system will depend on your test needs as outlined on subsequent pages and your resources. The table below is intended to give you an idea of the type of system you should most likely consider based on your resources and projected use.

Table 2

Type of System (see pages 40-44 for descriptions)	Level and Type of Resources Needed	Number of Users and Frequency of Use That Can Be Accommodated
1. File of tests	Low money, time, expertise, equipment	Few users and/or infrequent use
2. Card file of items	Low money, low-medium time, some expertise for item review, low equipment	Few users with frequent use <u>or</u> moderate numbers of users with infrequent use
3. Items stored on computer (use existing word processing programs)	Low money, low-medium time, some expertise for item review and equipment, requires a word processor	Few users with frequent use <u>or</u> moderate numbers of users with infrequent use
4. Item information stored on computer (use existing data base managers with possible development of some software)	Medium-high money, time, expertise, low-medium equipment	Many users and/or frequent use

Table 2 (continued)

5. Both items and item information on computer; features limited (use existing micro-computer-assisted packaged software	Money depends on available equipment, low-medium time, some expertise for item review and equipment	Few users with frequent use or moderate number of users with infrequent use
6. Sophisticated computerized systems Both items and item information on computer; features extensive (use mainframe computers)	High money, time, expertise, equipment	Many users and/or frequent use

GENERAL TYPES OF ITEM BANKING SYSTEMS

Type 1--File of Tests

The very simpliest type of item banking you can do involves gathering extant tests from as many sources as you can, going through them to note general content and grade levels, and storing them as is. Then when you develop a test, you can go back through and get ideas (or items) from these tests.

The advantages of this system are low cost, minimal start-up time, and no requirements for an elaborate categorizing scheme. The last point is important if you are producing tests on a nonregular basis for a variety of uses. In this case, any particular categorization scheme might not work.

The disadvantages come in producing tests from the bank. Search and production time is longer (although, usually, not as long as if you do not have the test collection), you might miss items if you get tired of going through the file of tests, you might not be systematic if you need to match certain local objectives, you probably will not have statistical information about items, and items might be of varying quality.

If the items comprising the tests are individually described using a more complex scheme than suggested above, the test file begins to take on the characteristics of a card file of items (Type 2).

Type 2--Card File of Items

A simple manual card file is effective for many users. Items can be gathered from any available free source--teacher-made tests, public-domain items (see a list of sources in Appendix A), item trading, etc. The manner in which items are classified depends on the user. The classification scheme could be developed by the users, or borrowed from another source. If the classification scheme is complex, (see page 46 for ideas on how items could be classified) an index should be prepared, cross-referencing items to the scheme. It's a good idea to number each item by major category (e.g., 5R10-010 for the tenth item which measures fifth grade reading objective number 10). Then items can be easily added to each category without renumbering. These numbers can then be cross-referenced to any categories by which items will be retrieved.

A more or less sophisticated classification scheme is decided upon first. Then items are collected, reviewed and stored by category. Items are added, revised and deleted as the bank is used. Standard instructions and item enhancements (graphs, reading passages, etc.) can also be stored. Items which require these enhancements would be marked. A cross-referencing index could be produced to simplify location of particular items.

Tests could be developed by pulling, lining up and photocopying the desired instructions and enhancements. Then items are numbered.

This system offers several advantages. It can be fairly inexpensive to set up (if you build as you go), it requires very little equipment, it is easy to use (if all users agree on the classification scheme and adhere to procedures for refiling items), and test production is relatively easy. It is especially good for small users (1 to 10) with few resources. It could handle fairly frequent use, and could be used to support classroom testing, minimum competency testing, diagnosis, mastery learning, survey testing and program evaluation (if norm referenced comparisons are not needed). Item use data can be added as you go along.

Disadvantages occur with many users, complex item classification, or need for auxiliary features. Many users could result in missing items, inadequate refiling of items and inadequate access to the bank. A complex item classification scheme (where items are classified along more than one dimension, e.g., content by difficulty by level of cognitive processing involved) can make searches time-consuming. Need for auxiliary functions such as test scoring, updating student skills records and cross-referencing items to instructional materials can make this option more awkward.

Type 3--Items Stored on Computer

A third fairly simple type of item bank is one in which items are classified and retrieved manually, but the items themselves are stored in a word processing file. In this situation, the user browses through an index or a hard copy version of the item bank; following selection, items are pulled from the word processing file, arranged on-line and printed. Item descriptors (e.g., sequence numbers, pointers to instructions or enhancements and item statistics--see page 46 for ideas on possible descriptors) might have to be stored in hard copy only, unless the word processor has a means for supressing some information when an item is retrieved. Instructions and reading passages could also be stored and inserted where they belong. The sophistication of this system really depends on the capabilities of the word processing software and printer. Sometimes codes and other item descriptors can be stored along with the items. This might facilitate a limited search for specific types of items.

The major advantage to this approach is that the final test often looks neater than when items are lined up and photocopied. Neatness can be an advantage when the item bank's task is to produce a camera-ready version for someone else to copy and use. Using a word processor also simplifies item revision and formatting.

The major disadvantages are typically due to limitations in the word processing software used. Visual displays (such as graphics) can rarely be stored. Thus, such enhancements must be stored separately and pasted into the final test later. Also, the software and printer might not include the characters, underlining, type sizes, type styles and spacing needed for some items (for example, those involving superscripts or chemical formulae). Another disadvantage is that it may be difficult for all users to have convenient access to the items, and to obtain training on the word processor. The system might work better if there were one central location that produced tests to specification. Finally, many word processors are not designed for efficient retrieval of information from large files. Many programs either

limit the size of a file or slow to a crawl when manipulating large files. Item selection can be tedious, therefore; you may need to scroll through the entire item bank document, marking the beginning and end of each desired item. Currently, the best solution to these problems is to establish procedures for marking the beginning and end of items and for coding important information about each item. A programmer can then write a program to pull items on the basis of key words, item numbers, or other characteristics. Organizing items into separate files by subject is also helpful because it keeps files small and thus cuts search time. Any of these solutions may require programming expertise.

Type 4--Item Information Computerized

In this approach, the classification scheme and other information about an item is kept on a computer; the items are kept in hard copy. (This is the opposite of Type 3, where the items were stored on the computer and the classifications were in hard copy.) This might be a good solution when there are a large number of items, each classified along a number of dimensions (e.g., content, difficulty and level of cognitive function involved, etc., as outlined on page 46. The computer could search the data base according to some specified set of criteria and list the identification codes of all, or a selected portion of items that meet the requirements. These items could then be pulled from a hard-copy file, lined up and photocopied (as with the card file system--Type 2 above). Final items could be selected, and computer information updated. This approach would be most useful for high volume banks in which tests are developed at a centralized location.

Since it is the information about the items which is computerized rather than the items themselves, auxiliary functions such as test scoring, cross-referencing items to instructional materials and routinely updating item use and statistics is easier. However, changing item text becomes harder than in Type 3 and overall visual quality of the tests developed suffers from the same problems as the card file (Type 2).

These systems do not necessarily have to be implemented on a large computer system. For example, several types of generalized software available on microcomputers could be utilized for item banking, e.g., programs for database management, spreadsheet analysis, statistics, test analysis, graphics, and communication. The uses and limitations of some of these are summarized in Appendix C. (See Estes and Deck, 1984 for a more detailed description of some of the uses and limitations of these programs for item banking.)

Type 5--Both Items and Descriptors Stored on Computer (Microcomputers)

A number of item banking software packages are available. Appendix B summarizes 34 of them. (A more complete description of these can be found in Deck and Estes, 1984 and Deck, Nickel, and Estes, 1985). While software is available for almost all types of microcomputers, capabilities tend to be fairly limited. Some software, for example, have only limited ways of adding, revising, deleting and retrieving items that are part of the software package. Other software provide no items (you enter your own) but offers greater editing and retrieving flexibility. Still other software emphasized scoring and recordkeeping. In general, Deck and Estes (1984 and Deck, Nickel,

and Estes, 1985) conclude that "The programs are disappointing. With only a few exceptions, these programs fail to make efficient use of the full capabilities of microcomputers.... Advanced programming techniques that would ensure snappy retrieval of items from large item files have not been used." This situation is changing, however, as evidenced by the most recent reviews (Deck, Nickel, and Estes, 1985).

The general functions these programs perform are listed below. If this type of item bank sounds useful, you should look at this list, identify the functions which are of most use to you, find out the capabilities of your machine, and review the package to make sure that it will do what you want it to. Remember that these programs are generally limited in capacity and that no microcomputer program will perform all the tasks listed below.

- Item bank available. Some item banking software comes complete with items, some does not.

- Various response formats. Some item banks only allow multiple-choice items.

- Classification scheme. Some item banks have very primitive classification schemes and do not allow you to classify items in your own way, or in multiple ways.

- Item storage capability. Many programs severely restrict the length and format of items.

- Creating and editing items. Most programs allow you to enter and edit items, but few allow full-screen editing (i.e., altering the text of an item as you would on a word processor--having the whole screen available for changes).

- Usage history. Few item banks provide ways to track item history or evaluate item quality through item statistics.

- Test printing. Most programs allow little control over test formatting or printing (e.g., superscripts and subscripts).

- Administration and scoring. Many of the programs support on-line test administration. (This might be also useful for drill and practice.) Some systems also support mark-sense readers.

- Student recordkeeping. A few programs allow automatic tracking of student scores. Usually, however, this tracking cannot be done by skills mastered, only by test scores.

- Computer testing. A few programs support student on-line testing.

Type 6--Both Items and Item Information Stored on the Computer (Mainframes)
These systems have specialized software, are designed with a particular application in mind, and have high volume use. Both classification schemes and items are usually stored in a computer. Three examples of such systems were described in the yellow pages of the last section.

Detailed advice on how to implement one of these systems is outside the scope of this handbook because so many interrelated decisions are necessary. However, important issues to consider when developing such a system (taken from Millman and Arter, 1984) are listed in Appendix D.

ITEM TYPES, NUMBERS AND CLASSIFICATIONS SCHEME

The use to which you will put your item bank has implications for the types of items you will have, how many you will need and how you will classify them.

Item Types

Item types include multiple-choice, matching, true-false, fill-in-the-blank, short answer, essay, and performance tasks. Many item bank developers use only multiple-choice items although there are instances of other types, including prompts for writing assessments. You may want to include other types for variety or because of student age, the cognitive level of the material you are testing and the subject matter. All types can be stored either manually or by computer. Multiple-choice, matching, true-false and fill-in-the-blank can be machine scored. If there is no particular reason to exclude item types, you might as well keep all you find. Add item type classifiers to your item descriptions.

Number of Items (Millman and Arter, 1984)

The ultimate number of items you will need depends on the frequency of testing, the extent to which students will be retested on the same content, the number of content areas to be covered and the amount of detail with which content areas must be covered. For example, to support a minimum competency testing program at three grade levels you will only need items which cover the minimum competencies at those grades. You probably will test only once or twice a year. You must be careful not to reuse any test item too frequently. You will not need a great range of item difficulty since you are only interested in determining whether a student "passes" or "fails."

Another example is an item bank to support classroom testing. Since your goal here is to assess levels and completeness of knowledge, you will need items that cover all the important parts of your curriculum at various levels of difficulty.

Rules of thumb that have been suggested for the number of items in an item bank are 10 items for each one that could be used on a testing occasion, and 50 items for each class hour of presented material (Prosser, 1974). In general, the more the better, unless extra items are of poor quality or make item retrieval and selection difficult.

Items do not have to be acquired all at once, although some applications require a good number of items before the bank can be used at all. (For example, a diagnostic system being offered to many on-line users.)

How Should Items Be Classified?

The classification scheme is the means by which items will be found once they are stored in the system. The best way to classify items is to use "descriptor words." Each item is assigned a number of words or codes that best describe it. When thinking about these codes, you should include everything that will be important when you want to find a particular item in the bank. Most frequently, items are coded by at least content and grade level--objective within subject within grade. It might also be useful for you to code each item on one or more of the following:

- Difficulty or other item statistics (such as latent trait calibrations)
- Response format (multiple-choice, true-false, matching, etc.)
- Source of the item
- Cognitive level (recall of facts, inference, application, etc.)
- Judged importance (essential knowledge, desirable knowledge, etc.)
- Cross-reference to curriculum materials
- Security level
- Readability level
- Previous use (number of times used, last use, groups tested)
- Content key words (situation portrayed in the item such as reading a recipe or reading an essay, or topic covered such as "oxygen").

The use to which you will put your item bank may affect how you will classify items. For example, if you will be supporting diagnostic testing, it would be nice to cross-reference items to curriculum materials, know the thinking processes involved in each item, and know item format. On the other hand, if you will be developing classroom tests using your bank, you might want your items cross-referenced to your textbook or objectives, and might also want some idea of the importance of the information covered.

If you do not have a scope and sequence or other way of classifying items by content, you can "borrow" the content classification scheme of some other source. Appendix A indicates some sources for classification schemes.

Things to Keep in Mind:

1. When setting up a computerized item bank you need to consider search capabilities of the software as you decide on the number of items, the classification scheme for items and how items will be stored. As the number of items and the number of codes increases, so does search time.

2. The more complex your system, the longer it will take to review and classify items. You need to balance ease of use with item information.

HOW WILL TESTS BE DEVELOPED FROM THE BANK?

When developing a test from an item bank you might need any or all of the following:

1. A way to pull items using criteria of interest.

2. A way to specify the number of items needed for each area.

3. Pointers to related items (e.g., items from the same reading passage), to support materials (e.g., reading passages or graphics that are stored apart from the item texts), to items which should never appear together, and to sets of instructions which should be used with particular items.

4. A way to examine and select or reject items, as well as to revise or write others.

5. A way to assess the whole test (e.g., in terms of overall content, difficulty, and item order).

6. A set procedure for producing the test once items are selected.

7. Procedures and policies governing who can use the bank.

Finding and Pulling Items

Your classification scheme determines how this will be done. If you have a manual system, you should set up your major filing categories by the most important criteria for selecting items--usually content/objective and grade level. This will facilitate browsing, which can be faster for routine searches than using an index. If you have a more complex way of classifying items, you might need to produce an index. A computerized system will almost always have a key word index by which to find items.

Number of Items

Before you develop a test from the bank you need to know how many items of each type you will need. For a computerized system you will need to specify this information to the computer if you want it to select the right number of items from memory for each objective. You need to decide whether the computer will randomly select items and whether it will select fixed numbers of items (e.g., 5) for each objective specified.

Pointers

Pointers can facilitate test production, although they are not in fact, essential. They do not have to be added before an item is used; they can be added as users notice important item characteristics, for example, two items that should never be used together because one item gives away the answer to the other. Several items which relate to the same display (e.g., a graph or reading passage) often will not be stored together because they deal with different topics or measure different skills. Pointers can serve to tell you

what reading passage or graph to pull for particular groups of items. Similarly, you might have stored different sets of instructions for various types of items in your bank. Pointers can assist in grouping all items which require identical instructions.

Item and Test Review

This is of most concern if the system is computerized or if users are not selecting items themselves. If a computer is selecting items for you, you need to be able to review and reject items, and tell the computer to find others. The computer will need to flag the rejected items so that they will not be selected again. If the item bank is not accessed directly by the user, it is advisable to have the user review items before the final test is printed. Since this procedure increases turnaround time, however, it must be balanced against testing frequency.

Test Production

Production refers to the actual way that the selected items are formatted, numbered and put on the finished test. In a manual system, the easiest way to do this is to keep items on cards which can be sorted, overlaid and photocopied. You need to keep fixed margins on the cards and keep all classifers and answers out of view. In a system where the items are stored by computer, there must be a way to sort the items in the manner desired, insert displays and instructions, and number items before they are printed.

In addition, you need a plan for producing an answer key before items are refiled.

Access to the Bank

Manual and computerized systems can be set up so that users select items for themselves or have others select items for them. Certain testing uses imply the type of access you will have. Infrequent testing for which all examinees will receive the same items (such as survey or minimum competency testing) imply tests that are developed from a central location. Classroom uses (routine classroom testing, diagnosis, and mastery learning) imply the need for direct user access to the bank. The latter uses usually require more review, direct user approval of items, and quick turnaround time. User access can be hard to coordinate when items are to be used by persons in more than one building. Multiple needs often prompt a move to a computerized system, district resources permitting.

AUXILIARY FUNCTIONS

Depending on the purposes for testing, an item bank might include "extras" to assist in testing and instruction. For example, if test results are to be used for diagnosis, it might be important to key individual items or groups of items to instructional materials. Other auxiliary functions are as follows:

Test Printing

Some systems will only enable you to get items. Others will enable you to format and print tests. The overlay and photocopy method works well in many cases and avoids many of the problems of on-line sorting, inserting instructions and displays, numbering and printing. In addition, few computer systems will store graphics so these will have to be inserted by hand later anyway. Also, many word processors have limited type characters which may make it difficult to store certain types of items (such as equations). Automatic test printing can be useful if items contain only standard text.

Scoring

You might want to have your tests computer scored regardless of whether the rest of the system is manual or computerized. This will require, however, at least a microcomputer and a test scanner. You might consider this option if scoring is extensive or particularly complex (as in some diagnostic schemes).

Reporting

Automatic reporting of results is usually associated with computer scoring and is usually done only for high volume item banks. Some microcomputer item banking software, however, will produce certain types of reports. (See Appendix B for more information about microcomputer software.)

Recordkeeping

Automatic recordkeeping is also associated with computer scoring and reporting. You might consider this option if you are using an application that requires frequent updating of student status such as mastery learning. In general, frequent diagnosis and assessment of student mastery calls for computer support. Some microcomputer software will do simple recordkeeping tasks; however, users at this level of sophistication typically have large computers and design their own software.

ACQUISITION AND MANAGEMENT OF ITEMS

At this point, you need to consider how items will be acquired, reviewed, and entered into the item bank, and how the items will be managed.

Item Acquisition

You have to decide how you will get items for your bank. Some considerations are listed below and relate to the general tradeoffs between item quality, expense and desired use.

It is almost always desirable to gather as many items as you can from other sources, rather than producing all needed items yourself. Items are available from commercial publishers, from publically supported institutions, and sometimes from individuals. Appendix A summarizes possible sources of items in various subject areas for various grade levels.

The major advantages in using existing item collections are: (a) It is usually less costly than writing your own; (b) It is less time comsuming; and (c) Items stand a chance of being better in quality. Even if your bank will be used only by a few persons, it can be very useful to obtain existing items. First, you will start out with more items than if you rely on your old tests or other local items. Second, having items from many sources can broaden your perspective on how to measure specific skills and knowledge. Third, the items sometimes have already been screened for quality and tried on students.

For small users, adding items as you go is usually the best procedure. Once you have decided the purpose of your item bank and its needed topics and levels, you can readily file items by topic and level as you find them. As the number of users, the applications for items and potential uses for results all increase, the formality with which items are selected, reviewed and categorized ahead of time must also increase. These factors largely determine the time and cost required for establishing an item bank. For example, for some uses, it is necessary to have large numbers of items available before the bank can be used at all. Such is the case, for example, with the three examples from the previous section--item banks used by a publisher, by a state department of education, and by a county.

Sometimes items need to be produced locally. This is the case if the subject matter is particularly idiosyncratic or if there is a strong local perception that items from other sources would never match local standards for quality, context or curriculum. Sometimes items are written to fill in "holes" that obtained items just don't cover. Item writing must become more rigorous as the importance of the test scores increases--for example in minimum competency testing. "Rigorous procedures" call for training item writers, reviewing items for content, bias and technical quality, and field testing. Estimates from test developers suggest that obtaining items for $10 apiece is less expensive than structuring a rigorous local item writing effort.

Small users, who will not be making critical, lasting educational decisions from their tests, need not be as rigorous. Items can be reviewed for problems as they are used.

Quality Control and Item Management

These questions are not unique to item banking. Whenever persons use test items there is a concern for quality. We mention these issues specifically here because one of the advantages in having an item bank is the chance to store and reuse good items and revise poor items. The idea is that the entire pool of items from which to choose will increase in quality over time. A good quality test item is one which measures the skill intended--no more, no less. To the extent that students can get the item right without knowing the right answer, or get the item wrong due to factors other than their knowledge, the item is poor.

As the importance of the test use increases (as with tests for promotion), the requirement for quality increases. Individual teachers can increase the quality of items in their test pools by noticing which items students often miss, and finding out why by asking them; by routinely keeping track of the number of students getting each item right; by looking at wrong responses to see if there is a possibility for confusion; and by having someone else take the tests and comment on the clarity and relevance of items. As items are used they can be upgraded in quality, as such analysis indicates a need.

For larger groups of users and where more formal systems are in place, there is need for more formal training in item writing, more careful review of items going in the pool, and more thorough examination of actual student performance on items to accommodate the various user's needs and preferences.

Technical reviews usually cover these concerns:

- Item clarity--Does the item have only one right answer? Does it ask only one question?

- Item bias--Would any particular group of examinees be more or less able to answer this question for reasons entirely apart from what they know? Item bias can be assessed statistically and/or more informally by representatives of important subpopulations.

- Technical quality--Can the question be answered without reading the stem? Does anything in the stem give away the answer? Does the item function well statistically, e.g., difficulty and discrimination?

Along with these technical concerns, you need to monitor the content of the items being used. Do the items measure the skills you want measured? Are the items classified properly? Again, small users can review these properties as items are used. Large users usually ensure proper classification and content up front through careful training and multiple reviews of classifications. Even if items with someone else's classifications are used, it is a good idea to review them for content because your idea of what is meant by a specific skill might not be the same as someone else's.

EXAMPLES OF ITEM BANKS

In order to clarify the various item banking options, the following yellow pages provide detailed examples of item banks that support various uses and involve various levels of computer usage.

Example 1--A Manual System Using A Card File

Item Acquisition and Classification

This item bank covers reading, language arts and math items in grades K-12. Items have been gathered over time from available sources and include--locally written and public-domain items as well as contributions from item-sharing banks. There are over 10,000 items in each subject areas; a large number are needed because of the many skills covered. Items are reviewed for quality before being placed in the item bank by testing specialists. Item use information is also available for revising items. A classification scheme was developed to cover all skills and subskills in the areas of interest. (See Appendix F for a portion of the reading classification scheme.) The classification scheme was developed through a literature review of skill hierarchies. Items are classified by assessment specialists before being placed into the bank.

Item Storage

Items are stored on 5x8 cards. One sample is shown on page 53. Classification codes refer to codes you will use to retrieve items later. Last use (optional) provides room to note test date and other relevant data so that the same students do not repeatedly get the same items. Pointers are used to identify other items which go with this one, items which should never go with this one (because one item gives away the other), sets of instruction which go with particular items, and/or other material not stored with the item itself (such as instructional materials). Visual displays associated with only one item are stored with that item. Visual displays and reading passages associated with more than one item are stored in a separate file. Items are laid out with standard margins so that they can be overlain and photocopied without exposing any of the identifying information.

Items are organized according to the way in which they will usually be selected for use (i.e., by skill level and subject). This makes browsing quick and efficient. Often it can be faster to identify desired items by browsing than by going to an index.

Item Management

As items are pulled for use, they are reviewed for quality and flagged for revision or deletion. Later, new items are added to the system as they are written to replace what was deleted.

Test Production

Desired items can be identified by looking through the index or by browsing. Promising items are pulled by ID number and examined. Unwanted items are replaced or rewritten. Pointers are checked for related items, items to be

avoided and instructions. The final items and instructions are laid out and photocopied. The test items are numbered. Last use information is updated on the cards, the answer key is prepared, and the items are refiled.

Auxiliary Functions
None.

Training/Equipment
No special training or equipment is needed.

Cost
Total cost for developing the classification scheme and entering the items is not available, but it is estimated that the equivalent of 1-2 person work years were involved.

Use
This bank supports infrequent test development from a central location. (It could, however, support higher frequency.) Users can identify the types of items they want or can browse through the file.

Leave enough room for spacing between tests

Item Identifer Code

Leave enough room for item #s

Item Text

FRONT

Use History

Classifiers	Date	N	Stats	Other Information
Grade				
Subject/Obj.				
Key Words				
Other _____				
Pointers				
Assoc. Items				
Items to Avoid				
Directions				
Visuals				
Other				
References				
Source				
Instructional Mats.				
Other Versions				

BACK

Example 2--A Prepackaged Microcomputer System

(Note: This description is taken from Deck and Estes, 1984.)

Item Acquisition and Classification

This particular item banking system comes with no items. The user must acquire and screen his/her own items using procedures independent of the program. Each item can be categorized on three dimensions--type of item (multiple-choice, true-false, etc.) and two user defined categories with six possible values each (e.g., six types of skills, cognitive function or difficulty). In addition, items can be described with key words.

Item Storage

Items in this system consist of the item text and an answer. About 700 items can be stored on a single disk, depending on the length of each item. The items are grouped in files which are limited by the memory size of the computer since the entire file must be loaded into memory before it can be manipulated. This feature ensures efficient editing but can lead to frustration if the user is careless about which items are entered in each file. The program does provide for merging and splitting item files. Item graphics must be stored in a hard-copy version and merged with items later.

Item Maintainance

The user enters and edits items just as they should appear on the page with many of the editing features found on most word processors. The program operates very fast on the item file in memory and pauses for only a short delay while the item file is loaded or saved.

Test Production

This system is intended for one user at a time. Once an item file has been loaded into memory, the user has many options for assembling the test: enter the item number, scan the items and flag desired items, select on three category schemes, or key words. An answer key and the master copy of the test are printed. The sort feature is used to create another form with items in another order. The test can be routed to a disk file for storage or even edited just before printing without affecting the stored item.

Auxiliary Functions

None.

Training

This package requires a TRS 80 Model III with at least one disk drive and 48K of memory. A new version is available which will run on the TRS 80 Model 4. The Radio Shack DMP series of printers or the Epson MX and FX series printers should be used with the package to utilize superscripting, subscripting and underlining.

This package is accompanied by a 77-page manual. The operation is well described even though no tutorial is included. This program is command- rather than menu-driven--users do not chose functions from a menu but control the program through commands which must be learned. This allows more flexibility in use but takes a little longer to learn.

Cost
This package costs about $200. Cost associated with item acquisition, entry, maintainance and test development varies with the user.

Use
This system is intended for a few users. It could support either centrally or user developed tests. Use could be fairly frequent.

Example 3
Item Classifications are Computerized, Items are in Hard Copy

Item Acquisition and Classification
About half the items were acquired from public domain sources (e.g., NAEP, old items, ERIC sources), and the other half were written locally. There are about 4,000 reading items, 4,000 math items, 4,000 language arts items, 1,500 life skills items, and 300-500 items in various vocational education topics. Items cover grades K-12. All items are multiple-choice.

Items are classified by identification numbers and a hierarchy of skills within categories. Other information available on each item includes Rasch calibrations and history of use.

Item Storage
Items are stored in binders along with graphics. The classification scheme is stored on a computer.

Item Management
Items are written, revised and updated as needed, usually at the request of a user. Some revision is done with each test developed from the bank. Items are reviewed for content by users, and for technical adequacy by bank operators using item statistics.

Test Production
Users specify the content areas and grade levels of items they want. Bank operators pull possible items and send them to users for review. After the items are approved, bank operators pull items from the notebooks, overlay and photocopy them, and send the clean copy to the user. The user is responsible for reproducing copies for use. Answer keys are automatically stored on the computer and are sent to users along with the test.

Auxiliary Functions
The tests can be machine scored using the answer key previously stored on the computer. Score reports in various formats can be provided.

Training and Equipment
Users have to be introduced to the classification scheme and procedures for accessing the system.

Cost

Development of the item bank began in 1977. The core (classification scheme, item acquisition and calibrations) was ready for use in nine months. Items are continually being added. Currently the bank uses about three professional FTE, two clerical persons and one programmer.

Use

The bank is maintained by a county office which develops tests for districts. Originally developed to assist with the production of high school minimum competency tests, the bank has expanded in use over time. About one test per day is now developed from the bank.

References

Deck, D.D. & Estes, G.D. Microcomputer Software Support for Item Banking. Paper presented at AERA, 1984.

Deck, D.D., Nickel, P.R. & Estes, G.D. Microcomputer Software Support for Item Banking. Northwest Regional Educational Laboratory, 300 S.W. Sixth Avenue, Portland, Oregon, 1985.

Estes G.D. & Arter, J.A. A Guide to Item Banking in Education, (Second Edition), Northwest Regional Educational Laboratory, 300 S.W. Sixth Avenue, Portland, Oregon, 1984.

Estes. G.D. & Arter, J.A. Item Banking for State and Local Test Development and Use, Northwest Regional Educational Laboratory, 300 S.W. Sixth Avenue, Portland, Oregon, 1983.

Ebel, R. Essentials of educational measurements. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1971.

Fall, 1984 issue of Educational Measurement - Issues in Practices.

Gronlund, N. Constructing achievement tests. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1982.

Gronlund, N. Measurement and evaluation in teaching. New York: MacMillan Publishing Company, 1976.

Mehrens, W. & Lehmann, I. Measurement and evaluation in education and psychology, New York: Holt, Rinehart & Winston, 1984.

Millman, J. & Arter, J.A. Issues in item banking. Journal of Educational Measurement, 1984, 21, 315-330.

Prosser, F. Repeatable tests. Educational Technology, 1973, 13, 34-35.

APPENDIX A

Item Bank Survey Summary

BEST COPY AVAILABLE

Name	Number of Items in:							Availability of:									
	Read-ing	Math	L.A.[1]	Sci.[2]	SS[3]	Life Skills[4]	Grades[5]	Items	Custom-ized Testing	Test Printing	Scoring	Reporting	Class. Scheme	Soft-Ware	Cross Ref.	State	RASCH
AK Objective and Item Bank	1050	1400[6]					K-8	X	X				X		X		
AIMS								X	X	X	X	X	X	X		X	X
Basic Skills Inventory	1500	1900[6]	1650			1000	K-12	X	X	X	X	X	X			X	X
Objectives Referenced Bank of Items/Tests	1000	2200[6]	550				K-12	X	X	X	X	X	X			X	
Sample Assmt. Exercises Manual Elem., Sec.	300	350[6]	200[8]			350	6-12	X	X				X			X	
New Brunswick Item Bank	Not Specified							X									
OAIP/BIMO	700	1250	32[8,9]	7500			6-12	X					X			X	
Knowledge Master	1200	1600[6]	6000	6000			6-12	X	X				X	X			
Delaware Item Bank	700	900[6]	250[9]				6-12	X	X	X		X	X				
Palm Beach Co. Item Bank		1500[6]					K-8	X					X		X	X	X
Life Role Skills Test Item Bank						1200	9-12	X					X		X	X	
HWEA	1100	1600	1200				3-12	X	X	X	X	X	X				X
Multiscore	1100	2400[6]	850	150	150	60	K-12	X	X	X	X	X	X			X	
IER Criterion Referenced Obj. and Item Banks	300	500[6,8]	400				K-8	X	X	X	X	X	X	X			X
Illinois Item Bank	50	150[6]					3-12	X					X			X	
Kansas Minimum Competency Item Bank	250	600[6]	225			190	K-College	X	X	X	X	X	X	X		X	X
Maryland Functional Testing	650	500[6]	47[8]			80	6-12		X	X	X	X	X	X	X	X	X
Michigan Ed. Assessment Program Tests	225	800[6]	Fine Arts 200				K-Adult	X	Only on Michigan				X		X	X	
SORT	600	800	80[8]	1400		800	K-12	X	X	X	X	X	X	X		X	
Common Body of Knowledge	1200	4800[6]	1600				K-12		Not Specified								
Montana Item Bank	200	50		150	Atti-tude 150	1350[11]	6,11	X	Only on Michigan								

	Item Classifiers:				Item Reviews:				
	Content	Diff.	Cog. Level	Response Type	Content	Technical	Bias	Field Test	Address
AK Objective and Item Bank	x				x		x	x	Alaska DOE Pouch F Juneau, AK 99811
AIMS	x	x	x		x	x	x	x	Merrill Test Division 1300 Alum Creek Dr. Box 508 Columbus, OH 43216
Basic Skills Inventory	x	x			x	x	x	x	Los Angeles County Schools 9300 Imperial Highway Downey, CA 90242
Objectives Referenced Bank of Items/Tests	x				x	x	x	x	CTB/McGraw Hill 2500 Garden Road Monterey, CA 93940
Sample Asmt. Exercises Manual Elem., Sec.	x		x		x	x	x	x	California DOE 721 Capitol Mall Sacramento, CA 95814
New Brunswick Item Bank	x				x	x		x	Dept. of Ed., King's Place Frederickton New Brunswick, E3B5H1
OAIP/BIMO	x	x		x	x	x		x	Ontario Ministry of Ed. 900 Bay Street Toronto, Ontario M7A1L2
Knowledge Master	Not Specified				x	x	x	Informal	Academic Hallmarks P. O. Box 998 85 Woodland (Suite B) Durango, CO 81301
Delaware Item Bank	x			x	x	x	x	Informal	Dept. of Public Instruction P. O. Box 1402 Dover, DE 19903
Palm Beach Company Item Bank	x	x			x	x			Palm Beach County Schools 3323 Belevedere Road West Palm Beach, IL 33402
Life Role Skills Test Item Bank	x	x		x	x	x	x	x	Laurens County Board of Ed. P. O. Box 2128 Dublin, GN 31031
NWEA	x	x			x	x		x	Northwest Evaluation Association 1410 S. 200th Street Seattle, WA 98148

BEST COPY AVAILABLE

	Item Classifiers:				Item Reviews:				
	Content	Diff.	Cog. Level	Res. Type	Content	Technical	Bias	Field Test	Address
Multiscore	x	x	x		x		x	x	Riverside Publishing Co. 8420 Bryn Mawr Avenue Chicago, IL 60631
IER Criterion Referenced Object. & Item Banks	x	x			x	x		x	Institute for Ed. Res. 793 N. Main Street Glen Ellyn, IL 60137
Illmuls Item Bank	Grade Level				x	x		x	Illinois State Board of Ed. 100 N. 1st Street Springfield, IL 62777
Kansas Minimum Competency Item Bank	x	x	x	x	x	x	x	x	University of Kansas School of Education Lawrence, KN 66045
Maryland Functional Testing	x			x	x	x	x	x	Maryland DOE 200 W. Baltimore Street Baltimore, MD 21201
Michigan Ed. Assmt. Program Tests	x			x	x	x	x	x	Michigan DOE P. O. Box 30008 Lansing, MI 48909
SORT	x	x		x	x	x	x	Informal	Saginaw School District 550 Millard Street Saginaw, MI 48607
Common Body of Knowledge	x		x	x	x	x			Jackson Municipal SD P. O. Box 2338 Jackson, MS 39225-2338
Montana Item Bank					x	x		Informal	Montana Office of Public Instr. Helena, MT 59620

[1]Language Arts including grammar, usage, mechanics, writing, outlining, spelling

[2]Natural Sciences

[3]Social Studies including Geography, History, Social Studies, Literature

[4]Career Development, Citizenship, Consumer Knowledge, Health, Voc. Ed., Basic Skills, Government, Cooperation

[5]Not all subjects necessarily at all grades specified

[6]Includes problem solving

[7]Test publisher

[8]Includes writing prompts

[9]Includes foreign language

[10]Includes computers

[11]Includes higher order thinking skills

[12]Includes affective items

APPENDIX B

Summary of Item Banking Software for Microcomputers

Table 1

SUMMARY OF ITEM BANKING PROGRAMS REVIEWED

(1985)

Program Title	Vendor	Computer	Major Features	List Price
AIMS	Charles Merrill Pub. Co. Test Division 1300 Alum Creek Dr. Box 508 Columbus, OH 43218	IBM PC-XT	Maintenance Test Assembly Item Bank	$900 min. (with items)
CREATE-A A-TEST	Cross Educational Software 1802 N. Trenton Ruston, LA 71270	Apple	Item Bank Maintenance Test Assembly	$ 90 $ 50/item bank
Exam Builder	A. U. Software P.O. Box 597 Colleyville, TX 76034	TRS-80 Commodore Apple IBM PC	Test Assembly Maintenance	$100
Exams and Examiner	Microsoftware Services P.O. Box 776 Harrisonburg, VA 22801	TRS-80	Maintenance Administration	$ 70 each $130 both
MicroCAT	Assessment Systems Corp. 2233 University Ave. Suite 310 St. Paul, MN 55114	IBM PC	Test Assembly Maintenance Administration	$975 and more
Multiple Choice Files	Compu-tations P.O. Box Troy, MI 48099	Apple Atari	Maintenance Administration	$ 30
P.D.Q. Builder	Micro Power and Light Company 12820 Hillcrest Road #219 Dallas,TX 75230	Apple	Maintenance Test assembly Administration	$ 45
Quiz Rite	Class 1 Systems 17909 Maple Street Lansing, IL 60438	Apple TRS-80 IBM-PC	Maintenance Test Assembly	$ 90
TAP	Addison-Wesley Pub. Co. Medical-Nursing Division 2727 Sand Hill Road Menlo Park, CA 94025	IBM PC	Item Bank Maintenance Test Assembly Administration	$125

Table 1 -- Cont'd

SUMMARY OF ITEM BANKING PROGRAMS REVIEWED

(1985)

Program Title	Vendor	Computer	Major Features	List Price
The Sage	Jagdstaffel Software 645 Brenda Lee Drive San Jose, CA 95123	Apple	Maintenance Administration	$ 62
Teacher Create Series (5 Programs)	Educational Courseware 67A Willard St. Hartford, CT 06105	Apple	Administration Maintenance	NA
TestBank	Holt, Rhinehart, & Winston 383 Madison Ave. New York, NY 10017	Apple	Item Bank Maintenance Test Assembly	NA
Test Rite	Class 1 Systems 17909 Maple Street Lansing, IL 60438	Apple TRS-80 IBM-PC	Maintenance Test assembly	$139
Tests Caicreat Caitake	EduSystems 3224 Lakeland Drive Berrien Springs, MI 49103	IBMPC	Maintenance Test assembly Administration	$ 75 Tests $100 All three
Tests Made Easy	Compu-tations P.O. Box 502 Troy, MI 48099	Apple	Test Assembly Maintenance	$ 30
Test-master	Midwest Software P.O. Box 214 Farmington, MI 48024	Apple CBM Commodore	Maintenance Test assembly	$ 35
Test-master Series	D.E.C. Computing 5307 Lynnwood Drive West Lafayette, IN 47906	IBM-PC Apple TRS-80	Item bank	$ 20/item bank
TestWorks	Milliken Publishing Co. 1100 Research Blvd. P.O. Box 21579 St. Louis, MO 63132	Apple	Maintenance Test Assembly Administration	$250

SUMMARY OF ITEM BANKING PROGRAMS REVIEWED (1984)

Program Title	Vendor	Computer	Potential Applications	Major Features	List Price
1. Classroom Administration Systems (CAS)	Micro Lab 2310 Skokie Valley Rd. Highland Park, IL 60035	Apple	Classroom testing	Maintenance Test assembly Scoring Recordkeeping	na
2. Examination	JAC Software 76 N. Western Avenue Elgin, IL 60120	Apple	Classroom testing Mastery learning	Maintenance Scoring	$100
3. Knowledge Master	Academic Hallmarks P.O. Box 998 Durango, CO 81301	Apple	Diagnosis Other	Test assembly Administration	$ 35 + $ 27 per item disk
4. Micro Test Administration System (MTAS)	S.R.A. 155 North Wacker Drive Chicago, IL 60606	Apple	Classroom testing Mastery learning	Maintenance Test assembly Scoring Recordkeeping	$295- 675
5. PACOR: Multiple Choice Tester	COMPress P.O. Box 102 Wentworth, NH 03282	Apple	Classroom testing	Maintenance Test assembly Recordkeeping	$ 50
6. Prism	The Psychological Corporation 7500 Old Oak Blvd. Cleveland, IL 60631	Apple TRS-80	Mastery learning Minimum competency	Test assembly Recordkeeping	$250- 675
7. Proctor	Comaldor P.O. Box 356 Postal Station O Toronto, Ontario M4A 2N9	Commodor	Classroom testing	Maintenance Test assembly Administration	$150
8. T.E.S.T	TYC Software 2128 W. Jefferson Rd. Pittsford, NY 14534	TRS-80	Classroom testing	Maintenance Test assembly Administration	$ 14- 25
9. Teacher Utilities Disk: Vol 1	MECC 2520 Broadway Drive St. Paul, MN 55113	Apple	Classroom testing Mastery learning	Maintenance Test assembly Administration	$ 37
10. Test Generator	COMPress (as above)	Apple	Classroom testing	Maintenance Test assembly	$ 50

SUMMARY OF ITEM BANKING PROGRAMS REVIEWED (1984)...continued

Program Title	Vendor	Computer	Potential Applications	Major Features	List Price
11. Test Maker	BERTAMAX, Inc. 3647 Stone Way North Seattle, WA 98103	Apple	Classroom testing	Maintenance Test assembly	$ 60
12. Testpac	Microdynamics Educational 2360 SW 170th Ave. Beaverton, OR 97006	Apple	Classroom testing	Maintenance Test assembly Administration	$ 35
13. Test Writer	Persimmon Software 502 C Savannah St. Greensboro, NC 27406	Apple	Classroom testing	Maintenance Test assembly	$ 35
14. The Learning System	Micro Lab (as above)	Apple	Classroom testing Mastery learning	Maintenance Administration	$ 75
15. The Test Bank	Advanced Technology Applications 4296 Tambor Court San Diego, CA 92124	TRS-80	Classroom testing Mastery learning Minimum competency Diagnosis	Maintenance Test assembly	$200

BEST COPY AVAILABLE

Table 2

Comparison of Features of Item Bank Software

	AIMS	Create-a-Test	Exam Builder	Exams& Examiner	MicroCAT	Multiple Choice Files	P. D. Q. Builder	Quiz Rite	TAP	The Sage
Item Bank Description										
May purchase item banks	X	X				X			X	
Use various response formats	X	X					X	X	X	
Link text passage to items	X				X				X	
Store classification scheme	X			X	X				X	
Allow more than 4 lines for item stem	X	X	X	X	X	X		X	X	X
Item Bank Maintenance										
Create and edit items	X	X	X	X	X	X	X	X	X	X
Use full-screen editing	X	X			X				X	
Store usage history									X	
Calculate item statistics					X					
Test Assembly										
Select by item number	X	X	X	X	X			X	X	
Select randomly	X		X	X		X	X			
Select by classification	X			X	X				X	
Print multiple forms	X								X	
Support special characters	X	X			X					
Administration and Scoring										
Administer test on-line				X	X	X	X			X
Support scanner										
Compute totals on test				X	X	X	X			X
Determine objective mastery					X					
Student Recordkeeping										
Use objectives approach										
Use gradebook approach				X	X		X			X
Use Factors										
Comprehensive manual	NA	X		X	X		X	NA	X	
Easy to Use	NA		X			X	X	X	X	
Reasonable performance	NA	X	X				X	X	X	X

BEST COPY AVAILABLE

Table 2 (Continued)

	Teacher Create	Test Bank	Test Site	Tests	Tests Made Easy	Test-master	Testmaster Series	TestWorks
Item Bank Description								
May purchase item banks		X						
Use various response formats		X	X	X	X		X	
Link text passage to items		X		X				
Store classification scheme		X	X	X			X	
Allow more than 4 lines for item stem		X	X	X	X	X		X
Item Bank Maintenance								
Create and edit items	X	X	X	X	X	X		X
Use full-screen editing		X		X		X		
Store usage history								X
Calculate item statistics								X
Test Assembly								
Select by item number		X	X	X	X	X		X
Select randomly		X		X	X	X		X
Select by classification		X	X	X			X	
Print multiple forms		X	X	X	X		X	
Support special characters		X						
Administration and Scoring								
Administer test on-line	X			X	X		X	
Support scanner								
Compute totals on test	X				X		X	
Determine objective mastery								X
Student Recordkeeping								
Use objectives approach								
Use gradebook approach								X
Use Factors								
Comprehensive manual	X	X	X	X		X	NA	X
Easy to use	X	X	X	X	X	X	NA	X
Reasonable performance		X	X		X	X	NA	X

BEST COPY AVAILABLE

COMPARISON OF FEATURES OF ITEM BANK SOFTWARE

Program Feature — Program:	1 CAS	2 Examination	3 Knowledge Master	4 MTAS	5 PACOR	6 Prism	7 Proctor	8 T.E.S.T.	9 Teacher Util.	10 Test Generator	11 Test Maker	12 Testpac	13 Test Writer	14 The Learn. Sys.	15 The Test Bank
ITEM BANK DESCRIPTION															
May purchase item banks			x	x		x						x			
Use various response formats	x	x	x				x	x	x			x	x	x	x
Link text passage to items					x	x								x	
Store classification scheme	x					x			x						x
Allow more than 4 lines for item stem				x								x			x
ITEM BANK MAINTENANCE															
Create and edit items	x	x	x	x	x		x	x	x	x	x	x	x	x	x
Use full screen editing															x
Store usage history	x						x							x	
Calculate item statistics	x	x		x											
TEST ASSEMBLY															
Select by item number	x			x	na		x			x		x			x
Select randomly			x		na		x	x	x	x			x		
Select by classification	x			x	na	x			x						x
Print multiple forms	x	x	x		na	x	x		x		x			x	x
Support underline, super-script, subscript, etc.															x
ADMINISTRATION AND SCORING															
Administer test on-line					x		x	x	x			x		x	
Support mark sense reader	x	x		x											
Compute totals on test	x	x		x	x		x	x						x	
Determine objective mastery		x													
STUDENT RECORDKEEPING															
Use objective approach						x									
Use gradebook approach	x			x	x										
USE FACTORS															
Comprehensive manual	x	na		x	na	x					na	x		x	x
East to use	na	na	x	x	na	x	na	na			na		na		x
Reasonable performance in file access, searches	na	na		x	na	x	na	na			na		na		x
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

x = program has feature
na = information not available

APPENDIX C

Summary of General Purpose Software for Microcomputers Which Could Be Used for Item Banking

Summary of Features of Generalized Software

Program Type	Potential Applications	Limitations
Word Processing	Enter and edit test items Format test for printing Use accessory programs for proofreading, checking readibility Preparation for typesetting	Difficult to select and retrieve items Files may be incompatible with other software
Database Management	Maintain item statistics, history of use Select items on item characteristics Store and retrieve item stems Maintain student recordkeeping Develop complete item banking system	Limited statistics available Files incompatible with other software Text handling limited or slow Programming experience needed for sophisticated applications
Spreadsheet	Maintain item statistics, history of use Select items on item characteristics	Must have database features (e.g. 1-2-3) May be awkward compared to database programs
Statistics	Evaluate test or item reliability and validity	Packages rarely include test analysis statistics Severe limits on number of items or cases
Test Analysis	Scan answer sheets with mark sense readers and score them Compute item statistics and test reliability	Only classical item statistics available Difficult to integrate with other software
Design Graphics	Produce drawings, figures, charts Create special symbols, formulas	Poor quality hardcopy output Difficult to store and retrieve by computer Slow printing, high storage demand
Communications	Transfer items or item data to another computer for further processing Access item bank maintained on larger computer	Cheap methods are slow Technical knowledge required to setup
CAI Authoring Systems	Integrate testing with on-line instruction Adaptive testing	Very labor intensive and expensive to develop

APPENDIX D

Item Bank Design Questions

From: Millman & Arter, 1984

Questions to be Answered in Designing Item Banking Systems

I. ITEMS

A. Acquisition and Development

1. Develop/use your own item collection or use collections of others?

   a. If develop your own item collection, what development procedures will be followed?

   b. If use collections of others, will the items be leased or purchased, and is the classification scheme sufficiently documented and the item format specifications sufficiently compatible for easy transfer and use?

2. What types of "items" will be permitted?

   a. Will open-ended (constructed response) items, opinion questions, instructional objectives, or descriptions of performance tasks be included in the bank?

   b. Will all the items be made to fit a common format (e.g., all multiple-choice with options a, b, c, and d ?

   c. Must the items be calibrated, validated, or otherwise carry additional information?

3. What will be the size of the item collection?

   a. How many items per objective/subtopic (collection depth)?

   b. How many different topics (collection breadth)?

4. What review, tryout and editing procedures will be used?

   a. Who will perform the review/editing?

   b. Will there be a field tryout, and if so, what statistics will be gathered, and what criteria will be used for inclusion into the bank?

B. Classification

1. How will the subject matter classifications be conducted?

   a. Will the classification by subject matter use fixed categories, keywords, or some combination of the two?

b. Who will be responsible for preparing the taxonomy?

c. How detailed will the taxonomy be? Will it be hierarchically or nonhierarchically aranged?

d. Who will assign classification indices to each item, and how will this assignment be verified?

2. What other assigned information about the items will be stored in the item bank? (See the attached list for potential attributes.)

3. What measured information about the items will be stored in the bank? (See the Appendix B list for potential measures.) How will the item measures be calculated?*

C. <u>Management</u>

1. Will provision be made for updating the classification scheme and items? If so:

a. Who will be permitted to make additions, deletions, and revisions?

b. What review procedures will be followed?

c. How will the changes be disseminated?

d. How will duplicate (or near duplicate) items be detected and eliminated?

e. When will a revision of an item be trivial enough that item statistics from a previous version can be aggregated with revisions from the current version?

f. Will item statistics be stored from each use, last use, or aggregated across uses?

2. How will items that require pictures, graphs, special characters, or other types of enhanced printing be handled?

3. How will items that must accompany other items, such as a series of questions about the same reading passage, be handled?

---

*This question is the subject of considerable controversy and discussion in the technical measurement literature. For example, to obtain a latent trait difficulty parameter, concern has been expressed about sample size, calibration procedure (Rasch, 3-parameter), linking models (major axis, least squares, maximum likelihood). and number of items common to the equating forms.

II. TESTS

A. Assembly

1. Must the test constructor specify the specific items to appear on the test or will the items be selected by the computer?

2. If the items are selected by the computer:

   a. How will one item out of several that matches the search specification be selected (randomly, time since last usage, frequency of previous use)?

   b. What happens if no item meets the search specifications?

   c. Will a test constructor have the option to reject a selected item, and if so, what will be the mechanism for doing so?

   d. What precautions will be taken to insure that examiners who are tested more than once do not receive the same items?

3. What item or test parameters can be specified for test assembly (item format restrictions, limits on difficulty levels, expected score distribution, expected test reliability, etc.)?

4. What assembly procedures will be available (options to multiple-choice items placed in random order, the test items placed in random order, different items on each test)?

5. Will the system print tests or just specify which items to use? If the former, how will the tests be printed or duplicated and where will the answers be displayed?

B. Administration, Scoring and Reporting

1. Will the system be capable of on-line test administration? If so:

   a. How will access be managed?

   b. Will test administration be adaptive, and if so, using what procedure?

2. Will the system provide for test scoring? If so:

   a. What scoring formula will be used (rights only, correction for guessing, partial credit for some answers, weighting by discrimination values)?

b. How will constructed responses be evaluated (off-line by the instructor, on-line/off-line by examiners comparing their answers to a key, on-line by computer with/without employing a spelling algorithm)?

3. Will the system provide for test reporting? If so:

a. What records will be kept (the tests themselves, individual student item responses, individual student test scores, school or other group scores) and for how long? Will new scores for individuals and groups supplement or replace old scores.

b. What reporting options (content/format) will be available?

c. To whom will the reports be sent?

C. Evaluation

1. Will reliability and validity data be collected? If so, what data will be collected by whom, and how will they be used?

2. Will norms be made available and, if so, based on what norm-referenced measures?

III. SYSTEM

A. Acquisition and Development

1. Who will be responsible for acquisition/development, given what resources, and operating under what constraints?

2. Will the system be made transportable to others? What levels and what degree of documentation will be available?

B. Software/Hardware Features

1. What aspects of the system will be computer assisted?

a. Where will the items be stored (computer, paper, card file)?

b. Will requests be filled using a batch, on-line, or manual mode?

2. Will items be stored as on large collection or will separate files be maintained for each user?

3. How will the item banking system be constructed (from scratch; by piecing together word processing, data-base management, and other general purpose programs; by adopting existing item banking systems)?

4. What specific equipment will be needed (for storage, retrieval, interactions with the system, etc.)?

5. How user and maintenance friendly will the equipment and support programs be?

6. Who will be responsible for equipment maintenance?

C. Monitoring and Training

1. What system features will be monitored (number of items per classification category, usage by user group, number of revisions until a user is satisfied, distribution of test lengths or other test characteristics, etc.)

2. Who will monitor the system, train users, and give support (initially, ongoing)?

3. How will information about changes in system procedures be disseminated?

D. Access and Security

1. Who will have access to the items and other information in the bank (authors/owners, teachers, students)? Who can request tests?

2. Will users have direct access to the system or must they go through an intermediary?

3. What procedures will be followed to secure the contents of the item bank (if they are to be secure)?

4. Where will the contents of the item bank be housed (centrally or will each user also have a copy)?

5. Who will have access to score reports?

IV. USE AND ACCEPTANCE

A. General

1. Who decides to what uses the item bank will be put? And will these uses be the ones that the test users need and want?

2. Who will develop the tests and who will be allowed to use the system? Will these people be acceptable to the examinees and recipients of the test information?

3. Will the system be able to handle the expected demand for use?

4. Will the output of the system likely to be used and used as intended?

5. How will user acceptance and item bank credibility be enhanced?

B. Instructional Improvement. If this is an intended use:

1. Will the item bank be part of a larger instructional/decision-making system?

2. Which textbooks, curriculum guidelines, and other materials, if any, will be keyed to the bank's items? Who will make that decision and how will the assignments be validated?

3. Will items be available for drill and practice as well as for testing?

4. Will information be available to users that will assist in the diagnosis of educational needs?

C. Adaptive Testing. If this is an option:

1. How will the scheduling of the test administrations take place?

2. How will the items be selected to insure testing efficiency yet maintain content representation and avoid duplication between successive test administrations?

3. What criteria will be used to terminate testing?

4. What scoring procedures will be followed?

D. Certification of Competence. If this is an intended use:

1. Will the item bank contain measures that cover all the important component skills of the competence being assessed?

2. How many attempts at passing the test will be allowed; when? How will these attempts be monitored?

E. Program/Curriculum Evaluation. If this is an intended use:

1. Will it be possible to implement the system so as to provide reliable measures of student achievement in a large number of specific performance areas?

2. Will the item bank contain measures that cover all the important stated objectives of the curriculum? That go beyond the stated objectives of the curriculum?

3. Will the item bank yield commensurable data that permit valid comparisions over time?

F. Testing and Reporting Requirements Imposed by External Agencies. If meeting these requirements is an intended use:

1. Will the system be able to handle requirements for program evaluation (e.g., Chapter 1), student selection into specially funded programs, assessing educational needs, and reporting?

2. Will the system be able to accomodate minor modifications in the testing and reporting requirements?

V. COSTS

A. Cost Feasibility

1. What are the (fixed, variable) costs (financial, time, space, equipment and supplies) to create and support the system?

2. Are these costs affordable?

B. Cost Comparisions

1. How do the item banking system costs compare to the present or other testing systems that achieve the same goals?

2. Do any expanded capabilities justify the extra cost? Are any restricted capabilities balanced by cost savings?

86A

2554e

APPENDIX E

Test Selection Forms

RATING SCALE FOR STANDARDIZED TESTS

Instructions: List the tests under consideration along the top of the chart below. Then respond to each question using the following codes:

2 - Good
1 - Fair
0 - Information not available
-1 - Weak
-2 - Unsatisfactory

Name of Test

A. VALIDITY (Use completed "Assessing Content Validity" chart to answer these questions.)			
1. Do the test items measure at least 75% of the objectives of the program? (There should be at least 3 items per objective)			
2. Do at least 50% of the test items directly measure the objectives of the program? (If no, stop rating that test.)			
3. Does the test reflect the relative emphases of the program?			
4. Is the test free of irrelevant features? (Such as reading level of directions and non-reading subtests; regional, cultural and sex biases; other irrelevant features)			
B. RELIABILITY			
1. Is the reliability of the test sufficiently high? ($r_{xx} \geq .85$)			
C. NORMS			
1. Are schools, school districts or cities of your size, geographic region and urbanism included in the norms sample?			

2 - Good
1 - Fair
0 - Information not available
-1 - Weak
-2 - Unsatisfactory

Name of Test

C. NORMS (continued)

2. Have norms been developed in a way which makes them representaive of the population they claim to represent?

3. Has the test been normed within the last ten years?

4. Do the empirical* norms dates correspond to the times when you intend to test?

D. TEST LEVELS

1. Are test levels sufficiently broad that the same level may be administered for both pre- and posttests?

2. Are test levels available at the functional level of all students?

E. TEST SCORES

1. Are scores reported as NCEs or percentile equivalents?

2. If out-of-level testing is contemplated, are expanded scale scores available?

3. Are other scores available which meet local district reporting needs?

*Note: Empirical norms dates are dates the publisher actually administered the tests to the standardization sample.

2 - Good
1 - Fair
0 - Information not available
-1 - Weak
-2 - Unsatisfactory

Name of Test

F. SCORING CONSIDERATIONS

1. Are the types of score reporting services available that the local district desires (e.g., individual reports, class reports)?

2. If out-of level testing is done, can the scoring service provide in-level percentiles?

3. If machine scoring is not used, can teachers hand score the tests and use the norms tables with a minimal degree of training?

G. ADMINISTRATION CONSIDERATIONS

1. Can teachers administer the tests with minimal training?

2. Does the amount of time required to administer the test (or subtests) meet the school schedule?

3. Is the method of administration (group vs. individual) appropriate?

H. USABILITY

1. Does the cost of the test fall within budgetary limitations?

2. Is the layout of the test (e.g., number of items per page, print size) appropriate?

3. Does the test fit in with other district needs?

4. Can test booklets be purchased separately for each subtest?

ASSESSING CONTENT VALIDITY

Write your program objectives, goals or expected outcomes in the left column. Write the names and levels of the tests you review in the blank columns (use separate columns for different levels of the same test battery). While reviewing each test, write the number of the test items which measure each objective, as well as those items which measure none of the program objectives.

Learning Outcomes (Objectives, Goals, etc.)	Tests Being Reviewed (Form and Level)				

APPENDIX F

Sample Classification Scheme For Reading Items

READING

A. COMPREHENSION AND ANALYSIS

1. Select the MAIN IDEA of a reading passage
2. Select the BEST TITLE for a reading passage
3. Identify SUPPORTING DETAILS in a reading passage
4. Identify the SEQUENCE OF EVENTS in a paragraph
5. Identify the CAUSE AND EFFECT relationship between elements in a paragraph
6. Select the correct CLASSIFICATION FOR a DESIGNATED FACT
7. Select the statement that correctly COMPARES what, when, where, why or how events happened
8. Select the statement that correctly CONTRASTS what, where, why, when or how events happened
9. Select the WORD to which a given REFERENT (pronoun, adjective or adverb) refers
10. Select the CONCLUSION given in a passage
11. Select a statement or FACT WHICH SUPPORTS the CONCLUSION
12. Select the EMOTIONAL SENSE to which the author is appealing
13. Select the PHYSICAL SENSE (e.g., sight, sound, taste, etc.) to which the author is appealing
14. IDENTIFY the FIGURE OF SPEECH in a passage
15. Identify the MEANING OF a FIGURE OF SPEECH
16. Select the statement which indicates that an INFERENCE can be drawn
17. Select meaning of GRAPHIC CLUE
18. Select the ORGANIZATIONAL PATTERN used by the author
19. Classify a statement as SPECIFIC OR A GENERALIZATION
20. Identify AUTHOR'S VIEWPOINT, bias or objectivity in a reading passage

A. COMPREHENSION AND ANALYSIS (continued)

21. Distinguish between FACT OR OPINION

22. Classify a passage as NON-FICTION OR FICTION

23. Select the PROPAGANDA TECHNIQUE used in a passage

24. Select the TONE OF a PASSAGE

25. Identify the AUTHOR'S PURPOSE for writing a passage

26. Select a statement which supports or refutes the AUTHOR'S CREDIBILITY

27. Judge the VALIDITY OF the AUTHOR'S CONCLUSIONS

28. Select a POSSIBLE SOLUTION to the problem presented in a passage

29. Select the most appropriate PREDICTION that can be made based on the passage

30. Identify EXPLICIT INFORMATION directly expressed in a reading passage

31. Identify the MEANING OF a WORD FROM its USE in a passage

32. Identify various aspects of CHARACTERIZATION (mood, changes, influencing factors, etc.)

33. Identify ELEMENTS OF FICTION

34. Identify use of LITERARY TERMS

35. Identify POETIC DEVICE (alliteration, onomatopoeia, assonance, consonance, rhyme scheme, stanza, etc.)

36. Distinguish FACT AND FANTASY

37. Identify RELEVANT/IRRELEVANT INFORMATION

38. Identify RELATIONSHIPS BETWEEN STATEMENTS (analogy)

39. Identify BIASED/UNBIASED INFORMATION

40. Identify CONNOTATIVE word MEANING

E. STUDY AND RESEARCH SKILLS

1. Identify the ALPHABETICAL ORDER of a given list of words.

2. Select the answer arrived at by FOLLOWING directions

3. Select the word that would be on the same page of a dictionary as two given GUIDE WORDS

4. Identify the sample word in a PRONUNCIATION GUIDE which illustrates the pronunciation of the vowel

5. Select the syllable with the primary ACCENT in a multi-syllable word based on the dictionary entry

6. Select the specified information from a TABLE OF CONTENTS

7. Select specified information from an INDEX

8. Identify how to locate a book in the CARD CATALOG (e.g., subject, author, title)

9. Identify the appropriate USE of common REFERENCE MATERIALS

10. Identify the BEST REFERENCE SOURCE for a given topic

11. Select best OUTLINE for a list of related phrases

12. Identify proper ORGANIZATION of LAIBRARY MATERIALS

13. Select the specified information from an APPENDIX

14. Select the specified information from a BIBLIOGRAPHY

15. Select the specified information from a COPYRIGHT

16. Select the specified information from a GLOSSARY

17. Select the specified information from an INTRODUCTION

18. Select the specified information from a DICTIONARY

F. VOCABULARY

1. Select a SYNONYM for a given word
2. Select a ANTONYM for a given word
3. Select the correct meaning of a HOMOGRAPH which has been used in a sentence
4. Select the correct HOMOPHONE to complete a sentence
5. Select the correct meaning for a MULTIPLE MEANING WORD which has been used in a sentence
6. Select the APPROPRIATE AFFIX for a root word to complete a sentence
7. Select the word with a PREFIX to match a given definition or complete a sentence
8. Select the word with a SUFFIX to match a given definition or complete a sentence
9. Select the correct MEANING OF a given WORD
10. Udebtuft appropriate WORD IN CONTEXT
11. Identify SIGHT WORDS

A. COMPREHENSION AND ANALYSIS (continued)

21. Distinguish between FACT OR OPINION

22. Classify a passage as NON-FICTION OR FICTION

23. Select the PROPAGANDA TECHNIQUE used in a passage

24. Select the TONE OF a PASSAGE

25. Identify the AUTHOR'S PURPOSE for writing a passage

26. Select a statement which supports or refutes the AUTHOR'S CREDIBILITY

27. Judge the VALIDITY OF the AUTHOR'S CONCLUSIONS

28. Select a POSSIBLE SOLUTION to the problem presented in a passage

29. Select the most appropriate PREDICTION that can be made based on the passage

30. Identify EXPLICIT INFORMATION directly expressed in a reading passage

31. Identify the MEANING OF a WORD FROM its USE in a passage

32. Identify various aspects of CHARACTERIZATION (mood, changes, influencing factors, etc.)

33. Identify ELEMENTS OF FICTION

34. Identify use of LITERARY TERMS

35. Identify POETIC DEVICE (alliteration, onomatopoeia, assonance, consonance, rhyme scheme, stanza, etc.)

36. Distinguish FACT AND FANTASY

37. Identify RELEVANT/IRRELEVANT INFORMATION

38. Identify RELATIONSHIPS BETWEEN STATEMENTS (analogy)

39. Identify BIASED/UNBIASED INFORMATION

40. Identify CONNOTATIVE word MEANING

E. STUDY AND RESEARCH SKILLS

1. Identify the ALPHABETICAL ORDER of a given list of words.
2. Select the answer arrived at by FOLLOWING directions
3. Select the word that would be on the same page of a dictionary as two given GUIDE WORDS
4. Identify the sample word in a PRONUNCIATION GUIDE which illustrates the pronunciation of the vowel
5. Select the syllable with the primary ACCENT in a multi-syllable word based on the dictionary entry
6. Select the specified information from a TABLE OF CONTENTS
7. Select specified information from an INDEX
8. Identify how to locate a book in the CARD CATALOG (e.g., subject, author, title)
9. Identify the appropriate USE of common REFERENCE MATERIALS
10. Identify the BEST REFERENCE SOURCE for a given topic
11. Select best OUTLINE for a list of related phrases
12. Identify proper ORGANIZATION of LAIBRARY MATERIALS
13. Select the specified information from an APPENDIX
14. Select the specified information from a BIBLIOGRAPHY
15. Select the specified information from a COPYRIGHT
16. Select the specified information from a GLOSSARY
17. Select the specified information from an INTRODUCTION
18. Select the specified information from a DICTIONARY

F. VOCABULARY

1. Select a SYNONYM for a given word
2. Select a ANTONYM for a given word
3. Select the correct meaning of a HOMOGRAPH which has been used in a sentence
4. Select the correct HOMOPHONE to complete a sentence
5. Select the correct meaning for a MULTIPLE MEANING WORD which has been used in a sentence
6. Select the APPROPRIATE AFFIX for a root word to complete a sentence
7. Select the word with a PREFIX to match a given definition or complete a sentence
8. Select the word with a SUFFIX to match a given definition or complete a sentence
9. Select the correct MEANING OF a given WORD
10. Udebtuft appropriate WORD IN CONTEXT
11. Identify SIGHT WORDS

BEST COPY AVAILABLE